Universidad Autónoma de Madrid

Departamento de Bioquímica

EVOLUTION OF EMBRYONIC PLURIPOTENCY

Beatriz Fernández-Tresguerres Torrecillas.

Madrid, 2011

Departamento de Bioquímica

Facultad de Medicina

Universidad Autónoma de Madrid

EVOLUTION OF EMBRYONIC PLURIPOTENCY

Beatriz Fernández-Tresguerres Torrecillas.

Licenciada en Ciencias Biológicas

Director de tesis: Miguel Manzanares Fourcade

CNIC, Madrid

Miguel Manzanares Fourcade, Investigador Senior del Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid.

CERTIFICA

Que Beatriz Fernández-Tresguerres Torrecillas, Licenciada en Ciencias Biológicas, ha realizado bajo mi dirección, en el Instituto de Investigaciones Biomédicas "Alberto Sols" (CSIC-UAM) de Madrid, y en el Centro Nacional de Investigaciones Cardiovasculares (CNIC), el trabajo titulado *Evolution of embryonic pluripotency*. El presente trabajo cumple con los requisitos necesarios para ser presentado y defendido como tesis doctoral.

Madrid, 14 de diciembre de 2010.

- Firma Director de tesis

Miguel Manzanares
Department of Cardiovascular Developmental Biology
Centro Nacional de Investigaciones Cardiovasculares-CNIC
Melchor Fernandez Almagro, 3
28029 Madrid, Spain

- Visto bueno Tutor

Pilar Santisteban Profesora Honoraria Bioquímica UAM

A mis padres

“ Nothing in Biology Make Sense Except in the Light of Evolution ”

Theodosius Dobzhansky, 1973.

AKNOWLEDGEMENTS

AGRADECIMIENTOS

En primer lugar GRACIAS con mayúsculas a Miguel Manzanares por todas las oportunidades que me has dado para crecer en lo profesional y en lo personal a lo largo de todos estos años. Por cómo me has empujado y animado asistir a reuniones, cursos, congresos o estancias. Por tus miles de ideas brillantes que compartes con nosotros con entusiasmo incansable que se contagia.

Gracias a mis compañeros, este trabajo es vuestro. A Susana Cañon, por acompañarme en mis primeros pasos en el mundo del pollo en particular y en la ciencia en general, a Teresa Rayón por los buenos momentos y su experta ayuda con los experimentos, a Eva Alonso por tu apoyo y cariño dentro y fuera del labo, muak! mil gracias. A los antiguos miembros Bárbara Pernaute y Miguel Crespo gracias por su entusiasmo e incansable ayuda y a los nuevos Cris Arias, Melisa Gómez y Luis Aguirre por su apoyo en la recta final.

Gracias a la gente del departamento, por todo! Se os va a echar de menos!. Gracias a Simon Bartlett por su excelentes comentarios a la tesis.

Thanks to Ian Chambers. Once I got used to your Scottish accent I learnt a lot from your expertise on our favourite gene. My stay in Edinburgh seems very far away now, but I will never forget all friends I shared my life with: Dougy, Nick, Bianca, Rodrigo, Violeta, Adam, Louise, Nicolas, Michaela and Stefan.

Thanks to Hisato Kondoh who gave me the opportunity to experience one of the most exciting periods of my life. You made me feel very comfortable in Osaka, thousands of kilometres away from home. Makiko Iwafuchi and “Taka”, I miss you a lot! You showed me a country full of surprises. Tatsuya Takemoto I smile every time I remember you, I really enjoyed working with you at the bench. Thanks to all members of the lab who help me when I was completely “lost in translation”. Makiko Kawata nobody would guess when we were in London that after years we would meet again in Japan!

Gracias a mis padres, sois increíbles. Siempre estáis ahí: aquí, en Edimburgo, en Japón o donde quiera que sea. Os admiro. Rafa y Paloma, simplemente geniales, como nos hemos divertido y lo que queda, ¡mejor imposible! Lunis, gracias por quererme todos los días como si no me hubieras visto en años.

Gracias a mi pequeña pero gran familia. Abuelito, siempre decías que era tu corresponsal por el mundo, sigo retrasmitiendo para ti desde aquí Pericles con toda la ilusión. Abuelita, tan risueña e incansable, la fuerza de todos, Belén, Rafael, Enrique, Violeta, Quique, Ceci, Ramón, Cristina, Ramonín, Violeta y Cayetana, ¡gracias!

¡Gracias chicas, Belén, Ana I., Sol, Inés, Mónica, Ana M. y Blanca! Después de tantos años de amistad siempre estoy deseando veros, gracias por todos esos buenos momentos que me han ayudado mucho.

Gracias Ana Fraile, como me has animado, juntas hemos pasado por cómicos momentos en la universidad, doctorado y viajes, imposibles de olvidar.

Javi, te he dejado para el final porque no he encontrado las palabras para decirte gracias, ni creo que las encuentre nunca. Lo eres todo para mí y más.

ABSTRACT

Pluripotency in the early mouse is established and maintained by a gene regulatory network under the control of a core set of transcription factors that includes *Oct4*, *Sox2* and *Nanog*. This network is shared by the embryonic stem cells which can give rise to any adult cell type thanks in part to this network, which shuts down the cells´ differentiation programs and keeps them in an undifferentiated state. While the network is largely conserved in eutherian mammals, very little information is available regarding its evolutionary conservation in other vertebrates. We have compared the embryonic pluripotency networks in mouse and chick by means of expression analysis in the pre-gastrulation chick embryo, genomic comparisons, functional assays of pluripotency-related regulatory elements in ES cells and blastocysts, and in vivo overexpression assays. We find that multiple components of the network are either novel to mammals or have acquired novel expression domains in early developmental stages of the mouse. We also find that the downstream action of the mouse core pluripotency factors is largely mediated by genomic sequence elements that are not conserved with chick. In the case of *Sox2* and *Fgf4*, we find that elements driving expression in embryonic pluripotent cells have evolved through a small number of nucleotide changes that created novel binding sites for core factors.

These findings suggest that the EP-GRN arose not only through the appearance of novel pluripotency genes but also by co-opting and duplicating existing genes and establishing new regulatory interactions between them. The *de novo* appearance of some of these interactions in mammals was confirmed by overexpressing Nanog in chick and mouse embryos. Furthermore, expression analysis of these embryos suggests additional conserved roles for *Nanog* as a repressor of neural and haematopoietic differentiation at gastrulation stages.

Our results show that the network in charge of embryonic pluripotency is an evolutionary novelty of mammals that may have evolved owing to the comparatively extended period during which mammalian embryonic cells need to be maintained in an undetermined state prior to differentiation. Further knowledge of how this embryonic pluripotency changed during evolution will provide a deeper understanding of its control and it will extend our ability to exploit the potential of stem cells.

La pluripotencia en el embrión temprano de ratón la establece y mantiene una red génica regulatoria bajo el control de un grupo central de factores de transcripción que incluye *Oct4*, *Sox2* y *Nanog*. Esta red la comparten las células madre embrionarias que pueden dar lugar a cualquier tipo de célula adulta gracias en parte a que esta red actúa bloqueando los programas celulares de diferenciación mientras mantiene a las células en un estado indiferenciado. Esta red está ampliamente conservada en mamíferos euterios pero poco se sabe de su conservación evolutiva en otros vertebrados. Hemos comparado las redes de pluripotencia embrionaria de ratón y pollo por medio de análisis de expresión del embrión de pollo pre-gastrulatorio, comparaciones genómicas, ensayos funcionales con los elementos reguladores relacionados con la pluripotencia en células madre y blastocistos, así como por ensayos de sobreexpresión in vivo. Hemos encontrado que múltiples componentes de la red son nuevos en mamíferos o han adquirido nuevos dominios de expresión en los estadios tempranos del desarrollo del ratón. También que la acción *downstream* de los factores centrales de pluripotencia en ratón, esta mediada por elementos de secuencia genómica que no están conservados en pollo. En el caso de *Sox2* y *Fgf4* encontramos que elementos que dirigen la expresión en las células embrionarias pluripotentes evolucionaron por un pequeño número de cambios en los nucleótidos que crearon nuevos sitios de unión para los factores centrales. Estos resultados sugieren que la red regulatoria génica de la pluripotencia embrionaria surgió no solo por la aparición de nuevos genes sino también por cooptación y duplicación de genes ya existentes y por la aparición de nuevas interacciones entre ellos. De hecho sobreexpresando Nanog en embriones de pollo y ratón confirmamos que algunas interacciones son nuevas en este último. Por otra parte, el análisis de expresión de estos embriones sugiere que *Nanog* tiene un papel adicional y conservado como represor de diferenciación neural y hematopoyética en estadíos gastrulatorios. Los resultados muestran que la red encargada de la pluripotencia embrionaria es una novedad evolutiva de mamíferos que surgió quizá porque éstos necesitaban de un mantenimiento del estado indeterminado de las células embrionarias mucho más prolongado que otros vertebrados. Un mayor conocimiento de cómo la pluripotencia embrionaria ha evolucionado nos ayudará a comprender en profundidad su control y a entender y usar todo el potencial de las células madre.

INDEX

ABBREVIATIONS

AO…….. Area opaca

AP…… Area pellucida

bdC…… Blastoderm-derived cells

BMPs…… Bone morphogenetic proteins

BS…… Binding site

CD……C-terminal domain

cDNA…… Complementary DNA

chr…… Chromosomes

Col1A1 …… Collagen type 1 alpha 1 locus

D…….. Dorsal

Dr…… *Danio Rerio*

EG……Embryonic germ

EGK…… Eyal-Giladi & Kochav

EPC ……Ectoplacental cone

EP-GRN…….. Embryonic pluripotency gene regulatory network

EPI…….. Epiblast

EpiS…….. Epiblast stem

EST…….. Expressed sequence tag

ExE …… Extraembryonic ectoderm

Gg…….. *Gallus gallus*

GRN…….. Gene regulatory network

HD…… Homedomain

HH…… Hamburger and Hamilton

HMG…… High mobility group

ICM…….. Inner cell mass

iPs…….. Induced pluripotent stem cells

IRES……. Internal Ribosome Entry Site

LIF…… Leukemia inhibitory factor

logFC…… Logarithm fold change

M2-rtTA…….. Tetracycline transactivator

Md…….. *Monodelphis domestica*

MHC…….. Major histocompatibility complex

MI…… Mesoderm invagination

Mm…… *Mus musculus*

ND…….. N-terminal domain

Oct4-DE…… *Oct4* distal enhancer element

pA …….. Polyadenylation signal

PaE…… Parietal endoderm

PE…….. Primitive endoderm

Pect…….. Primitive ectoderm

PGC…….. Primordial germ cells

POU…….. Pit Oct Unc

qPCR…… Quantitative PCR

SA …….. Splice acceptor

TE…….. Trophectoderm

tet…….. Tetracycline

TetOP…….. Tetracycline operator

TF…… Transcription factor

TP …….. Totipotent precursor

TS …….. Trophoblast stem

V…….. Ventral

VE…….. Visceral endoderm

WR…….. Tryptophan repeat

XEN…….. Extraembryonic endoderm

Xt…….. *Xenopus tropicalis*

ZGA…….. Zygotic genome activation

INTRODUCTION

1. EARLY MOUSE DEVELOPMENT

Embryonic development is the process by which all lineages and cell types that form an organism from a single cell are specified. For a short period during the embryonic development of mammals, a subset of cells retains the capacity to give rise to all embryonic lineages including the germline (Smith, 2005). This unique capacity, called pluripotency, is lost progressively as cells differentiate and is preserved only in the primordial germ lineage, the precursors of the gametes, which retain pluripotency through later development and into adulthood. Pluripotency is also exhibited by embryonic stem cells in culture. Pluripotency is established and maintained by a gene regulatory network in which a core set of transcription factors controls an extended network of other genes. Despite the importance of this network, little is known about how it arose during evolution.

1.1 Pre-implantation development

Over the first four days after fertilization, the mouse embryo passes through a series of stages before implanting in the lining of the uterus. These early embryonic stages in mammals are quite distinct from early stages in non-mammals, where embryos rely on the nutritional deposits carried in the egg. In mammals, embryo nourishment depends exclusively on the mother, so the embryo must implant in order to survive and grow. Therefore the early differentiation events that occur during this period are devoted to the formation of supporting structures to enable stable implantation and nutrition, and to the segregation of embryonic lineages.

The fertilized egg divides symmetrically to give rise to an embryo of two cells and subsequent divisions form an embryo of four and eight cells (Fig. 1). Transition from maternal to zygotic transcripts begins during these first divisions and is known as the zygotic genome activation (ZGA) (Cockburn and Rossant, 2010).

After 2.5 days (E2.5), the embryo undergoes compaction and the cells acquire a polarized epithelial morphology that will be maintained in the outer cells during the subsequent divisions of 16 and 32 cells (late morula). The distinction between inner and outer cells gives rise at 3.5 days to the first lineage decision of the embryo: the

formation of the inner cell mass (ICM) and the surrounding trophectoderm (TE). At the same time fluid is pumped into the embryo and a cavity known as blastocoele is formed. With the formation of the blastocoele, the mouse embryo is now considered a blastocyst. Twenty four hours after the formation of this structure, in the second lineage decision, the ICM gives rise to the epiblast (EPI), which will generate the embryo proper, and the primitive endoderm (PE), that will generate an extraembryonic tissue, the yolk sac (Fig. 1). Other extraembryonic tissues that will mostly constitute the placenta will come from the TE.

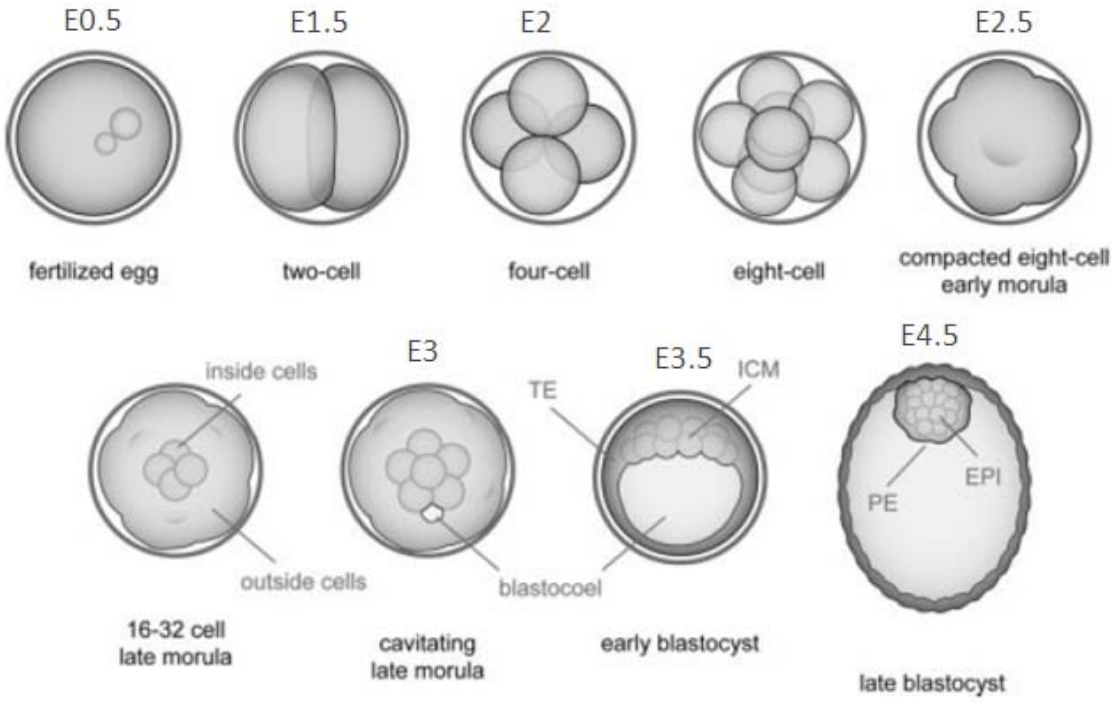


Figure 1. Stages of preimplantation mouse development from fertilization to the formation of the blastocyst. The fertilized egg undergoes three rounds of cleavage division to reach the eight cell stage. In the next stage, named the early morula, blastomeres undergo subsequent divisions and a process of compaction and polarization at 2.5 days. In the late morula this process generates differences between the inside and outside cells that later will give rise to the ICM and TE lineages respectively. The outer cells form the trophectodermal epithelium and pump fluid internally to form the blastocoele cavity at 3 days, which continues its expansion through the early blastocyst (E3.5) and late blastocyst (E4.5). In this late stage, the ICM segregates into two lineages, the epiblast (EPI) and the primitive endoderm (PE). *Modified from* (Yamanaka *et al.*, 2006).

1.2 Post-implantation development

By day 4.5 of development the embryo has implanted in the maternal uterine wall through the polar trophoectoderm, which directly overlies the ICM (Fig. 2A). Later these cells proliferate and expand to form the extraembryonic ectoderm (ExE) and the ectoplacental cone (EPC). Trophectodermal cells away from the ICM form the mural trophoectoderm; these cells stop dividing but continue to endoreduplicate their DNA to differentiate into the trophoblast giant cells (Fig. 2B) (Rossant and Cross, 2001).

Between days 4.5 and 6, cells in the epiblast proliferate to reach the distal pole of the blastocyst, aided by the formation of the ExE. This phenomenon reduces the size of the blastocoele cavity and results in the invagination of the primitive endoderm, which remains as a layer around the epiblast and the trophoectoderm (Fig. 2B). Later, primitive endoderm cells differentiate. Some PE cells form a single layer with the trophoblast giant cells, called the parietal endoderm, which will give rise to the yolk sac. In contrast, the primitive endoderm next to the epiblast and trophoblast forms the visceral endoderm. By 6.5 days the anteroposterior axis of the embryo is established when the anterior visceral endoderm migrates and the mesoderm is specified as the posterior region (Fig. 2C). This is the onset of gastrulation, a process that will lead to the formation of the three germ layers of the embryo: the ectoderm, mesoderm and definitive endoderm.

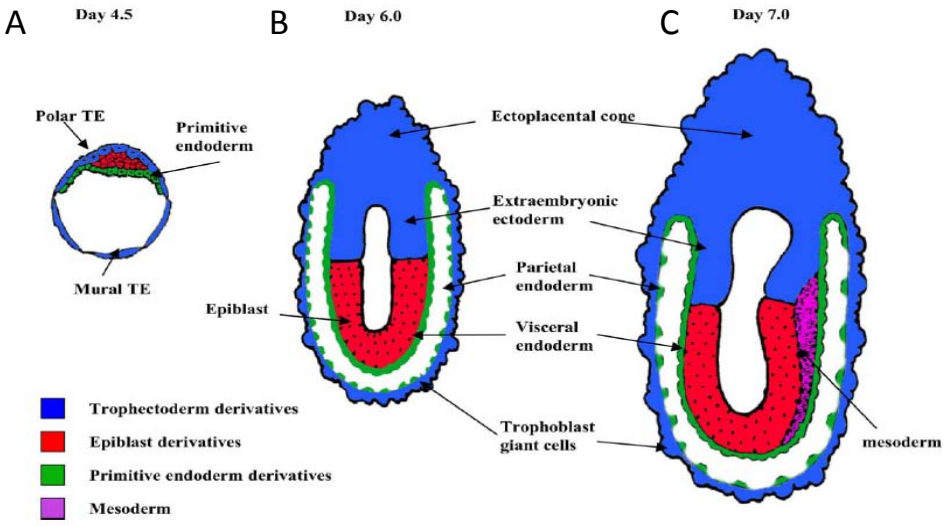


Figure 2. Lineage descendants of the three lineages of the blastocyst up to the time of gastrulation (Rossant, 2004).

1.3 Blastocyst derived stem cell

The succession of cells from blastocyst to gastrula is in a transient state that is lost with the differentiation events that take place during development. However, this transient state can be "captured" by removing cells at different stages and culturing them *in vitro*. Thus in the early blastocyst, the ICM is the source of embryonic stem cells (ES) and trophectoderm the source of trophoblast stem cells (TS), whereas in the late blastocyst the epiblast is the source of epiblast stem cells (EpiS) and the primitive endoderm is the source of XEN cells. At later stages the primordial germ cells (PGCs), precursors of the gametes, are the source of embryonic germ cells (EG) (Fig. 3).

All these cell types have the ability of self-renewal, which confers them with the capacity to proliferate indefinitely when cultured under defined conditions. Furthermore, they have the potential to differentiate into any cell type of the tissues for which they are precursors. Consequently cells derived from ICM, epiblast or primordial germ cells can give rise to all lineages of the embryo and are therefore considered pluripotent. Moreover, the derived cells in culture constitute a representative *in vitro* model for the study of the gene regulatory logic that controls animal embryonic development.

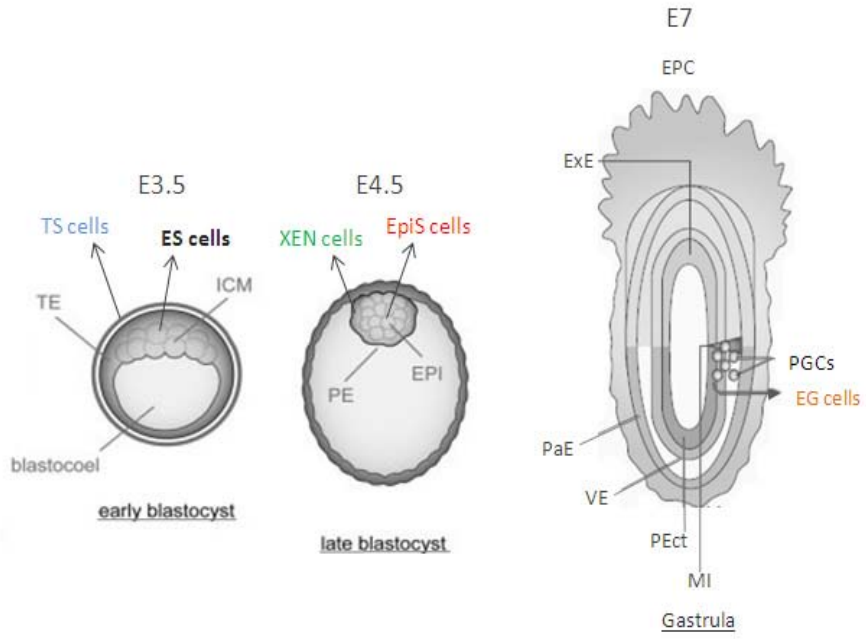


Figure 3. Embryonic pluripotent lineages as source of stem cells. In the early blastocyst, trophoblast (TS) and embryonic stem cells (ES) can be derived from TE and ICM, respectively. In the late blastocyst, the extraembryonic endoderm (primitive endoderm and its derivates) and epiblast are the sources of XEN cells and EpiS cells, respectively. At subsequent stages, the ability to derive these cells from the mouse embryo is progressively lost. The embryo starts gastrulation, which involves the formation of the three germ layers and the specification of the primordial germ cells (PGCs) in the epiblast from day 6.25; PGCs are the source of embryonic germ cells (EG). TE, trophoectoderm; ICM, inner cell mass; PE, primitive endoderm; EPI, epiblast; XEN, endoderm cells; EPC, ectoplacental cone; ExE, extraembryonic ectoderm; PaE, parietal endoderm; VE, visceral endoderm; PEct, primitive ectoderm; MI, mesoderm invagination. *Modified from* (Boiani and Scholer, 2005; Yamanaka *et al.*, 2006).

2. GENETIC CONTROL OF BLASTOCYST LINEAGES

2.1. Gene regulatory networks

Underlying the control of animal development we find thousands of genes that are arranged in gene regulatory networks (GRNs) that specify the sets of genes that must be expressed in specific spatial and temporal patterns. This progressively generates developmental patterns and executes the construction of multiple states of differentiation. The control system consists of modular DNA sequences; each module is activated or repressed by regulatory proteins such as transcription factors (TFs). TFs recognize specific DNA sequences in associated genes, controlling their transcription.

Some regulatory proteins control the activities of genes encoding other regulatory proteins, and their functional linkages then define the core networks (Davidson and Levin, 2005).

Current understanding of the GRNs underlying pre-implantation stages of mouse development is based mainly on studies with mutant mice and derived blastocyst cells. The results show that the segregation of lineages in mouse development is highly regulated by GRNs in which different transcription factors are expressed specifically in each cell type.

2.2. ICM versus TE fate

Several transcription factors are known to play key roles in determining the ICM and TE fates. *Cdx2* is required for TE development, while *Oct4* (official symbol *Pou5f1*), *Nanog* and *Sox2* are involved in establishing the ICM. In the mouse, *Cdx2* is first detected at the eight-cell stage and becomes restricted in the late morula to the outer cells, the future TE (Niwa *et al.*, 2005). *Oct4*, *Nanog* and *Sox2* are initially expressed in a stochastic fashion (Dietrich and Hiiragi, 2007) and only later become restricted to the inside cells, the future ICM cells. Reciprocal repression between *Cdx2* and the *Oct4-Nanog-Sox2* axis reinforces the segregation of the ICM and TE fates, together with the autoregulation of *Cdx2* and *Oct4*. As a result, upregulation of *Cdx2* in the outside cells is followed by downregulation in the same cells of *Oct4*, *Nanog* and *Sox2* (Fig. 4) (Cockburn and Rossant, 2010).

Null mutants have been obtained for *Cdx2* and *Oct4*, and both exhibit defects that correlate with their expression patterns. *Cdx2* mutant embryos form blastocysts but the TE in these embryos loses its epithelial integrity because *Oct4* and *Nanog* are ectopically expressed in outside cells at the blastocyst stage. Thus, the embryos cannot differentiate further, eventually leading to blastocyst death before implantation (Strumpf *et al.*, 2005). *Oct4* null embryos form a functional TE, and the blastocysts are able to implant; however, they die soon after because they lack ICM derivates such as the epiblast and yolk sac (Nichols *et al.*, 1998).

2.3. Epiblast versus primitive endoderm fate

The second lineage decision, between epiblast and primitive endoderm, depends on the activity of the transcription factors *Nanog* and *Gata4/6.* The ICM is composed of two cell types arranged in a mosaic pattern: the presumptive PE cells express *Gata6* and *Gata4* and the EPI cells express *Nanog*. *Nanog* is required for EPI determination and later signals from these *Nanog* expressing cells are essential for the formation of the PE (Fig. 4) (Messerschmidt and Kemler).

Nanog-deficient embryos die before implantation; the *Nanog*-null ICM cells lose their differentiation potential and are only capable of TE differentiation or undergo apoptosis (Silva *et al.*, 2009). *Gata4/6* mutants show deficits in establishing PE and fail to maintain PE derivates (Soudais *et al.*, 1995).

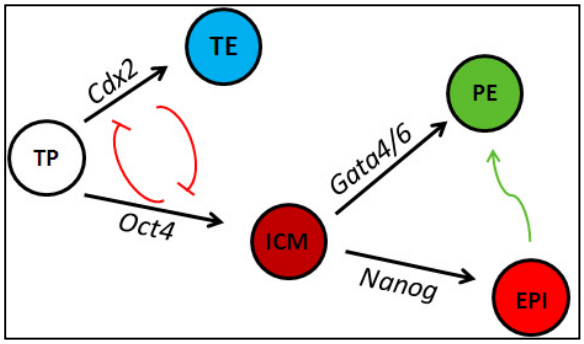


Figure 4. Lineage decisions in the pre-implantation mouse embryo. Mutually antagonistic *Cdx2* and *Oct4* drive the first lineage decision. In the second lineage decision, *Nanog* specifies EPI fate and promotes PE formation, possibly by a combination of paracrine signalling and cell-cell interaction. TP, totipotent precursor. *Modified from* (Messerschmidt and Kemler, 2010).

2.4. Peri-implantation stages

Axis formation and patterning, which occurs between 5.5 and 6.5 days, is driven by the expression and reciprocal interactions of other signalling molecules. These molecules are also necessary for the maintenance and proliferation of the trophoblast.

Fgf4 and *Nodal*, which are first triggers of differentiation at these stages, are expressed in the epiblast*, Fgf4* is directly downstream of *Oct4* and *Sox2* (Yuan *et al.*, 1995). *Fgf4* is widely expressed during early pre-implantation development, and then becomes confined to the ICM at the blastocyst stage and is eventually restricted to the early post-implantation epiblast. *Fgf4* has a role in PE formation (Feldman *et al.*, 1995) and

in the expression patterns of other FGF pathway components in the early embryo (Arman *et al.*, 1998; Chai *et al.*, 1998). *Nodal*, which is also found at early post-implantation stages, promotes posterior cell fate in the epiblast and signaling to the visceral endoderm, all leading to the establishment of anterior-posterior polarity (Brennan *et al.*, 2001).

Both genes also have a role in the extraembryonic lineage through their inductive action on the extraembryonic ectoderm, contributing to the expression of *Eomes, Cdx2* and *Esrrβ* transcription factors and preventing precocious differentiation of the ExE into the ectoplacental cone (Fig. 5) (Niswander and Martin, 1992; Rappolee *et al.*, 1994). *Eomes,* which is downstream of *Cdx2*, is initially expressed in the TE at the blastocyst stage and is restricted to the ExE after implantation (Ciruna and Rossant, 1999). *Esrrβ* transcripts are first detected in the ExE at E5.5 (Luo *et al.*, 1997).

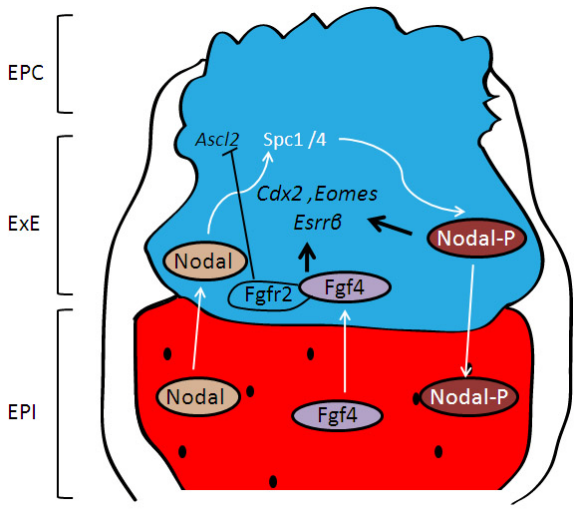


Figure 5. Reciprocal inductive interactions between the ExE and EPI. Nodal is processed in the ExE by the convertases Spc1 and Spc4. Processed Nodal (Nodal-P) maintains the expression of *Cdx2, Esrrβ* and *Eomes* in the ExE and contributes to the establishment of the embryo axis in the EPI. *Fgf4*, together with its receptor in the ExE, *Fgfr2,* also contributes to the expression of *Cdx2, Eomes* and *EsrrB,* and represses expression of the differentiation promoter *Ascl2*, restricting it to the EPC. ECP, ectoplacental cone; ExE, Extraembryonic ectoderm; EPI, Epiblast. (Arman *et al.*, 1998; Guzman-Ayala *et al.*, 2004)

3. PLURIPOTENCY

3.1 ES cells and the embryonic pluripotency gene regulatory network (EP-GRN)

ES cells possess many features that make them an attractive model for studying the molecular basis underlying embryonic development. ES cells are an invaluable tool for the study of mouse gene functions thanks to their ability to contribute to chimeras (Tam and Rossant, 2003), and can also be modified genetically or cultured under controlled conditions *in vitro* to monitor effects of the intra or extracellular environment.

Studies in ES cells have begun to unravel the GRNs that govern embryonic pluripotency. Furthermore, it is easy to obtain enough material from ES cell culture to perform genome-wide DNA binding protein localization studies, such as chromatin immunoprecipitation followed by microarray (ChIP-chip) or by direct sequencing (ChIP-seq) (Chen *et al.*, 2008; Kim *et al.*, 2008). These studies have increased our understanding of the molecular basis that controls pluripotency and have generated extensive datasets for further analysis.

Research with mouse ES cells has shown that pluripotency in ES cells and in the blastocyst results from the expression of a small network of transcription factors whose core members are the products of the *Oct4*, *Nanog* and *Sox2* genes (Fig. 6) (Silva and Smith, 2008). These genes actively maintain the undetermined state while also repressing the differentiation programme (Boiani and Scholer, 2005; Niwa, 2007). The core factors act together through auto- and cross-regulatory interactions and also through direct and overlapping binding to multiple locations throughout the genome, where they regulate downstream target genes (Fig. 7) (Boyer *et al.*, 2005; Loh *et al.*, 2006; Marson *et al.*, 2008). These transcription factors lie at the core of an extended GRN, which includes other transcription factors as well as epigenetic modifiers and signalling molecules. Their coordinated activity is ultimately responsible for the pluripotent state by controlling gene transcription in a combinatorial manner. In fact the differential regulation of the target genes is determined by the extent of regulatory element occupancy by diverse and multiple factors (Kim *et al.*, 2008).

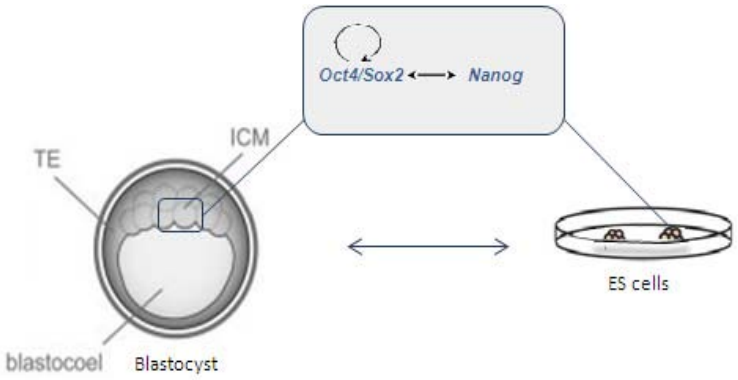


Figure 6. Pluripotency results from the action of a core set of transcription factors in the ICM cells and in ES cells. The core members of the network *Oct4*, *Sox2* and *Nanog* control an extended network of genes. ES cells offer a frozen state in development, in which the GRN responsible for embryonic pluripotency (EP-GRN) is maintained indefinitely, and the activation of differentiation programmes can be controlled experimentally by culture conditions (Silva and Smith, 2008).

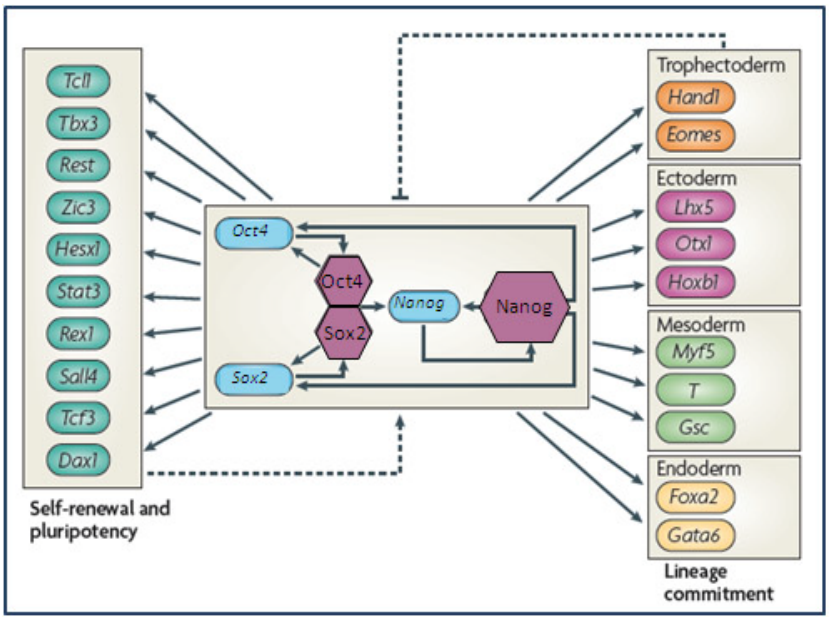


Figure 7. The core pluripotency GRN in ES cells. *Oct4*, *Sox2* and *Nanog* are central to the maintenance of the ES cell undetermined state. These genes co-occupy many target genes, activating other ES cell-specific genes while repressing the expression of genes associated with lineage commitment. The commitment to specific cell lineages results in changes in gene expression that include downregulation of the core transcription factors and the upregulation of differentiation genes. These three transcription factors regulate their own and each other's expression in a highly coordinated manner. This is achieved through multiple positive protein-protein and protein-DNA feedback interactions. For example, Oct4 and Sox2 form a heterodimer that positively regulates the expression of *Oct4*, *Sox2* and *Nanog* (Boyer *et al.*, 2005; Loh *et al.*, 2006). In the core network blue shapes denote genes and pink hexagons denote the corresponding proteins. Dotted arrows indicate potential feedback mechanisms from downstream targets, back to the core circuit. *Modified from* (Macarthur *et al.*, 2009).

Pluripotency is maintained through a complex interplay between the core network and peripheral signalling molecules. In the embryo these molecules are provided by the extracellular environment, whereas in ES cells extrinsic factors are added to the growth medium to create an appropriate external environment. For example, the cytokine leukemia inhibitory factor (LIF) and serum containing bone morphogenetic proteins (BMPs) are required to suppress differentiation and activate genes involved in self-renewal (Williams *et al.*, 1988; Ying *et al.*, 2003). Alternatively, a cocktail of enzyme inhibitors can be added to the medium to block differentiation signals generated autonomously by ES cells (Silva and Smith, 2008).

The central role of the core factors in pluripotency became more evident with generation of induced pluripotent stem cells (iPs), first obtained in 2006 (Takahashi and Yamanaka, 2006). Since then it has been shown that a wide range of murine and human somatic cells can be reprogrammed back to pluripotent stem cells by exogenous expression of specific combinations of transcription factors. The specific factors needed for reprogramming depend on the type of somatic cells being reprogrammed. Nonetheless, these combinations always include *Oct4* and *Sox2*, which are alone sufficient to initiate the reprogramming cascade (Takahashi and Yamanaka, 2006). Depending on the cell type, *Nanog* and other factors are included, such as *Klf-4*, *Lin28* or *c-Myc* (Judson *et al.*, 2009; Takahashi and Yamanaka, 2006; Yu *et al.*, 2007). After the introduction of the combination of transcription factors, the somatic cells reach the undifferentiated state by passing through a series of reprogramming events that include the demethylation and posterior activation of *Oct4*, *Sox2* and *Nanog*, which will be consistently expressed in the iPs.

ES cells share many features with EG cells, which are derived from primordial germ cells (PGCs) (Fig. 3). From about day 7.5 onwards, expression of the core factors *Nanog* and *Oct4* becomes restricted to PGCs, where they play critical roles in germ cell specification and development (Chambers *et al.*, 2007; Okamura *et al.*, 2008). EG derived cells thus express the core pluripotency transcription factors, and despite their distinct origins are in a similar molecular ground state to ES cells (Leitch *et al.*).

3.2. Biochemical and functional properties of EP-GRN core factors

The functional form of the Nanog protein is a homodimer. Dimerization requires a tryptophan repeat region located in the C-terminal half of the protein (Fig. 8A). In culture, constitutive expression of mouse *Nanog* is able to maintain the ES pluripotency state in the absence of LIF (Chambers *et al.*, 2003). This capacity is lost when the tryptophan repeat-region is deleted or mutated (Mullin *et al.*, 2008; Wang *et al.*, 2008) because Nanog dimerization is required for its ability to interact with other pluripotency proteins and therefore to promote self-renewal and repress differentiation (Fig. 8B) (Torres and Watt, 2008; Wang *et al.*, 2008).

Experiments in ES cells have shown the importance for pluripotency maintenance of a tight regulation of the core factors. Analysis of an *Oct4*-null mouse ES cell line, in which self-renewal is sustained by a tetracycline-suppressible *Oct4* transgene, revealed that suppression of the transgene triggered differentiation into trophectoderm (Kunath *et al.*, 2004). Surprisingly, when the same transgene was expressed in heterozygous ES cells (*Oct4*$^{+/-}$), the additional amounts of Oct4 protein did not lead to better maintenance of pluripotency but caused differentiation (Niwa *et al.*, 2000). In contrast to *Oct4*, *Nanog* deficiency does not commit ES cells to differentiation, and *Nanog*-null ES cells continue to self-renew even though they have an increased propensity to differentiate. Thus, while *Nanog* is indispensable for the formation of the ICM and germ cells, in ES cells it appears to act as a rheostat, conferring variable resistance to differentiation. *Nanog* levels fluctuate and cells with reduced *Nanog* levels are more prone to differentiate. These low levels would be like a “window of opportunity”, in which intrinsic or environmental perturbations can be consolidated into a lineage commitment decision or may be overcome by re-expression of *Nanog* (Chambers *et al.*, 2007; Kalmar *et al.*, 2009).

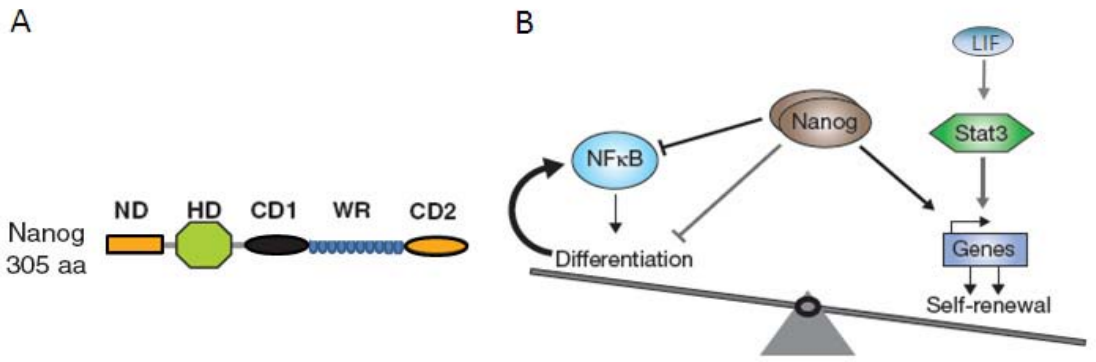


Figure 8. A) Mouse Nanog protein domains. The N-terminal half of Nanog contains a serine-rich motif in the N-terminal domain (ND) and a DNA-binding homedomain (HD). In the C-terminal half, two C-terminal domains, CD1 and CD2, are separated by a dimerization domain (WR), in which every fifth residue is a tryptophan; mouse Nanog has 10 copies of this motif (Chambers and Tomlinson, 2009). **B) Model showing the mechanism of action of Nanog in ES cells.** Nanog inhibits the expression of NFκB-dependent genes required for differentiation and synergistically activates Stat3-dependent promoters, bypassing the requirement of LIF .*Modified from* (Torres and Watt, 2008).

4. EP-GRNs IN NON-MAMMALIAN VERTEBRATES

Certain features of early lineage determination in the mammalian blastocyst can be considered evolutionary novelties, in the sense that they do not appear in other vertebrates. For example, non-mammalian vertebrates have no homologue of the TE, and the placenta is clearly a mammalian innovation. This raises the question of the extent to which genetic control of these early phases of mammalian development, in particular the establishment and maintenance of embryonic pluripotency, is conserved in other vertebrates.

Current knowledge is very much limited to the mouse, mainly due to the unique resource represented by ES cells. Little is known about the EP-GRNs present in other organisms, even though pluripotency is a necessary and transient stage in the development of any multicellular organism that passes through an obligatory one-cell stage as part of its life cycle. A few very recent studies have examined the evolutionary conservation of EP-GRNs in ES cells and early embryos between three

mammalian species: human, mouse and cow (Kunarso *et al.*, 2010; Xie *et al.*, 2010). These studies show that the EP-GRN varies significantly between mammals, but this variation appears to occur mainly in peripheral components of the network. Although interesting, these comparisons provide little insight into how deeply the EP-GRN is conserved beyond mammals and how it appeared during evolution.

Published reports in non-mammalian vertebrates are limited to the description and analysis of homologues of the core mammalian EP transcription factors *Oct4* and *Nanog*. *Oct4* homologues have been described in two species of fish. Early zebrafish express *Pou2/spg* and medaka embryos express *Ol-Oct4*. *Ol-Oct4* is also expressed in adult gonads of medaka; however, this pattern is not seen for zebrafish *pou2/spg* (Burgess *et al.*, 2002; Sanchez-Sanchez *et al.*, 2010). A medaka homologue of *Nanog* has been identified and has been linked to proliferation during early embryo development (Camp *et al.*, 2009).

Three tandem duplicated *Oct4* homologues have been isolated in the anuran amphibian *Xenopus*. One of them, *Xlpou91,* has the capacity to maintain pluripotency in murine *Oct4* null ES cells, and thus shares functional similarity with mammalian *Oct4* (Morrison and Brickman, 2006). No homologues of mammalian *Nanog* have been found in *Xenopus*. In contrast, axolotl, a urodele amphibian, contains homologues of *Oct4* and *Nanog*: *Axoct-4* is expressed in the early embryo but not in the PGCs (Bachvarova *et al.*, 2004), and *AxNanog* can sustain pluripotency in mouse ES cells (Dixon *et al.*, 2010). In chick, the *Oct4* and *Nanog* homologues also sustain mouse ES cell function in culture (Canon *et al.*, 2006; Lavial *et al.*, 2007).

While these reports point to a functional conservation of these factors among vertebrates, some territories of early expression are not conserved in non-mammals (Canon *et al.*, 2006) and the wiring of the network in non mammals is not known. Therefore we lack a clear understanding of the evolution of EP-GRN, and there is a need to compare the mouse EP-GRN with the regulation at equivalent stages in non-mammals.

4.1 The chick as a model

Birds, like mammals are amniotes: they lack a larval stage and instead they have a terrestrially adapted egg protected by a specific arrangement of extraembryonic membranes.

The chick (*Gallus gallus*) is an ideal model for making comparisons with mammals. The lineages of birds and mammals are separated by huge time gap of around 310 million years. This places them at a sufficient evolutionary distance for important differences to have arisen, while being sufficiently close to allow meaningful comparisons. In contrast, the early development of marsupials does not differ markedly from eutherians (Selwood and Johnson, 2006), and the other mammalian group (the monotremes) is not so easily available (Fig. 11).

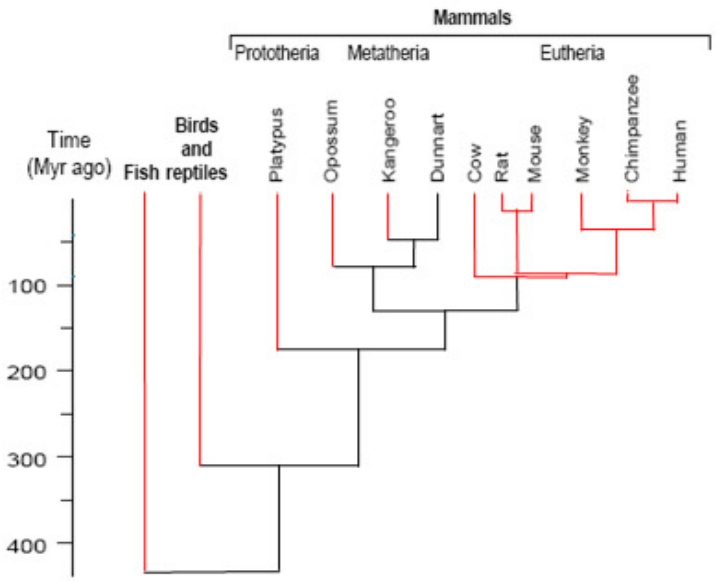


Figure 11. Relationships of the major mammalian groups with other vertebrates. The time scale is in millions of years (Myr). The ancestors of birds and mammals diverged around 310 million years ago. Marsupials comprise a single taxonomic unit (Metatheria), traditionally considered the sister group of Eutheria from which they diverged 130 million years ago. Red lines represent groups from which whole genome sequence are available. *Modified from* (Graves and Westerman, 2002).

Furthermore the recent sequencing of the chick genome (consortium, 2004) has confirmed the chick as a very valuable resource for comparative genomic analysis. Several characteristics make it especially useful for the identification of conserved non-

coding regions: the mentioned evolutionary position filling a gap in the genomes sequenced so far; the compactness of its genome, 40% respect to human and mouse; and that the chick genome has not undergone any recent duplication like teleosts fishes or many anuran amphibians. Hence this last characteristic means that in most cases there is 1:1 correspondence between homologous genes in mammals and birds, which includes a high level of conservation in intronic and flanking non coding regions likely to contain important regulatory elements. In fact it is estimated that at least 70 megabases (Mb) in the chick genome sequence are likely to encode functional, conserved elements, and there are long blocks of conserved synteny between chick and mammal genomes and a low rate of chromosome translocations (Stern, 2005).

Moreover the chick embryo is a classical embryological model because of its ready availability. Methods for transient transgenesis are available and allow efficient alteration of gene function during early embryonic development. Electroporation, the most popular method of transient transfection, has allowed studies of gene gain or loss of function (Sauka-Spengler and Barembaum, 2008). What is more, blastoderm-derived cells can now be derived in vitro from pre-primitive streak stages under specific culture conditions. These cells have been claimed to be the chick equivalent of mouse ES cells and might open the way to understanding the control of pluripotency in non-mammalian species (Petitte *et al.*, 2004).

4.2 Chick pre-gastrulation development

Fertilization in the chick occurs during the time between release of the oocyte and its entry into the oviduct. Peristaltic movements carry the fertilized egg down the oviduct, a journey that takes about 20-23 h and during which the egg becomes surrounded by albumen (egg white), egg membranes and the shell.

Early embryonic development also occurs during this time: in the hen's oviduct the fertilized cell on top of the yolk undergoes cleavage divisions that result in the formation of a disc of cells called the blastodisc or blastoderm (Fig. 12) (Fig. 13 A, B). Between 12 and 14 hours after fertilization, the anterior-posterior axis is specified by gravitation.

The early embryonic development is arbitrarily divided into 14 stages, numbered in Roman numerals (stages I to XIV) according to Eyal-Giladi & Kochav (EGK) (Eyal-Giladi and Kochav, 1976). The blastodisc or blastoderm corresponds to stage EGK-X, which is already composed of 20000-50000 cells, called blastodermal cells. Zygotic genome activation occurs at this stage (Zagris *et al.*, 1998).

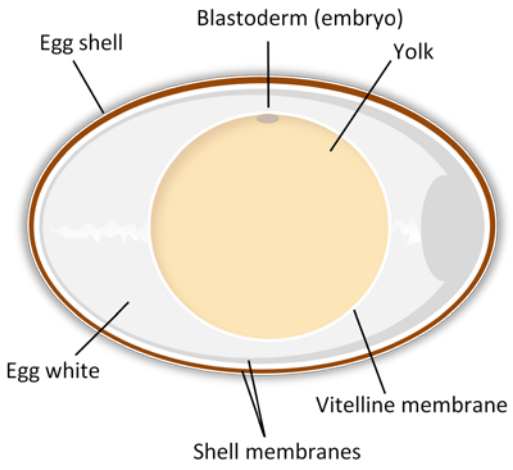
By the time the egg is laid (EGK IX-X) the embryo can be divided morphologically into a darker area called the area opaca (AO), which is the peripheral part and attached to the yolk, and a more translucent area called the area pellucida (AP), in the central part of the embryo overlaying the subgerminal space, (Fig. 13C). The AO gives rise to extraembryonic structures, while the entire embryo and some of the extraembryonic tissues develop from the AP. Some AP cells delaminate to establish a layer of cells overlaying the yolk beneath the epiblast in the AP. This layer, called the hypoblast, gives rise to the extraembryonic endoderm. Hypoblast cells derive from two sources: the ingression of cells from the epiblast and the posterior marginal zone, a band of epithelial cells at the posterior lateral boundary between the AO and AP (Fig. 13D).

Later the epiblast cells undergo a number of morphogenetic movements that enable the onset of gastrulation with the establishment of the primitive streak, the future axis of the embryo (Fig. 13E).

Embryonic development after laying is arbitrarily divided into 46 stages denoted with Arabic numerals (stages 1-46) according to Hamburger and Hamilton (HH) (Hamburger and Hamilton, 1992), where HH stages 1-3 align with EGK stages XII- XIV.

Figure 12. Fertilized hen's egg when laid.

The fertilized egg is confined to a small patch, several millimetres in diameter, lying on a large mass of yolk. Cell cleavage in the oviduct results in the formation of a disc of cells called the blastodisc or blastoderm *Modified from* (Wolpert, 2002).



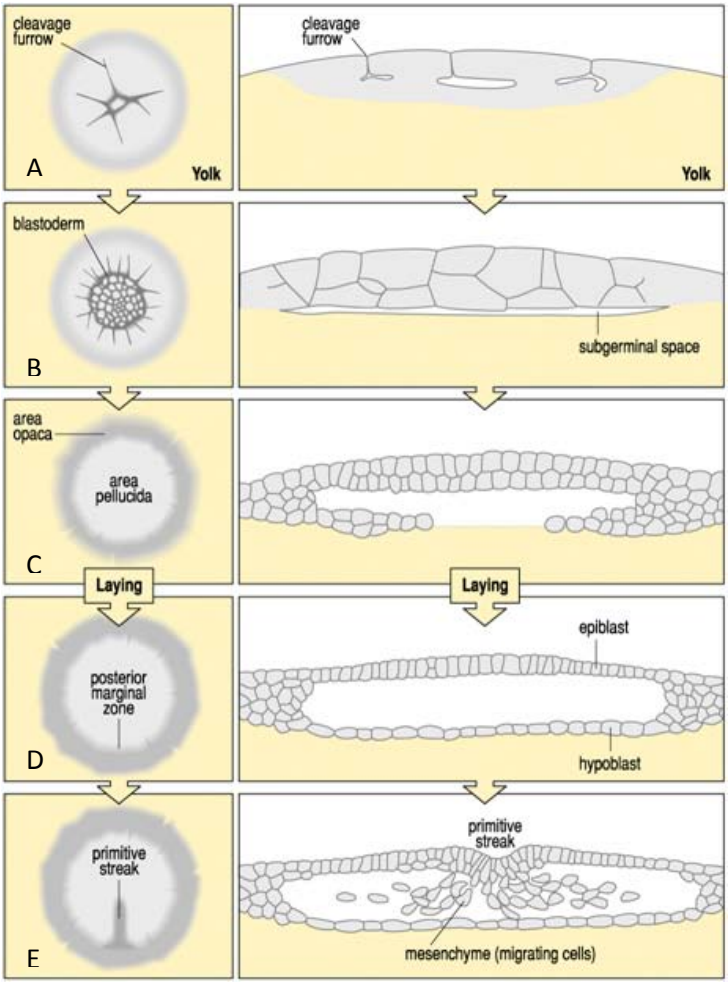


Figure 13. Pre-gastrulation development in the chick embryo. A) Before the egg is laid early cleavage furrows extend downward from the surface of the cytoplasm, but the cleavage is incomplete and initially does not separate the blastoderm from the yolk. **B)** Later cleavage results in a circular blastoderm several cells thick. **C)** The thinner central area, which overlays the subgerminal space, is known as the area pellucida, whereas the thicker marginal region is the area opaca. **D)** In the cavity a layer of cells develops over the yolk, called the hypoblast. This gives rise to the extraembryonic endoderm, while the epiblast will give rise to the embryo proper. **E)** The primitive streak starts to form and gradually extends further anteriorly. The primitive streak appears as a dark double line where a large number of epiblast mesenchymal cells have piled up, waiting to ingress (Wolpert, 2002).

4.3 Chick orthologues of mouse early specification genes

Unlike the mouse, where *Nanog* expression is found in the epiblast and thereafter becomes restricted to PGCs (Yamaguchi *et al.*, 2005), the chick orthologue is not expressed in the whole epiblast at early stages but is restricted to individual PGCs scattered over the epiblast (Canon *et al.*, 2006).

Similarly to the situation in mouse, an *Oct4*-related gene is ubiquitously expressed in pre-gastrulation stages in both AO and AP and later becomes restricted to the epiblast when the primitive streak starts to form. At later stages its expression appears in the PGCs, a feature also shared with its murine counterpart (Lavial *et al.*, 2007).

Cdx2 and *Eomes*, which are critical for the specification and development of the trophectoderm in mouse, are expressed in the chick embryo in the AO at pre-gastrulation stages, similarly to their expression in the TE of the mouse blastocyst. However, these genes are expressed in the reverse order to that seen in mouse TE (where *Eomes* is expressed downstream of *Cdx2*). Unexpectedly, chick *Eomes* is also present in a novel domain, the PGCs, which does not occur in mouse (Pernaute *et al.*, 2010). The *Fgfr2* orthologue is not expressed at pre-gastrulation stages in chick (Pernaute *et al.*, 2010), while in mouse it is expressed in the TE and its derivative, the extraembryonic ectoderm (Rossant and Cross, 2001).

4.4 Comparing chick and mouse pre-gastrulation stages

There are several key differences between early mouse and chick development before gastrulation. The most obvious is the existence of a nutritive yolk that feeds the embryo, a structure absent from mammalian embryos, where nourishment must be obtained through the TE-derived placenta. However, like the mouse, one of the earliest events to take place in the chick embryo is the separation of extraembryonic and embryonic territories. This segregation occurs shortly before laying (EGK-IX-X) with the distinction of the AO and AP. Soon after, the primary hypoblast arises from delamination of cells from the AP. This arrangement is topologically similar to that in mouse, with first the separation of extraembryonic and embryonic lineages and the subsequent formation of the primitive endoderm (Fig. 13) (O'Farrell *et al.*, 2004).

A particularly interesting stage for comparison with mouse embryos would be the pre-streak stage (EGK X-XIII), when the hypoblast has already delaminated from the epiblast but gastrulation has not yet begun. This stage has been proposed to align more precisely than any other with the late mouse blastocyst (4.5-5.5 days) (Fig. 13).

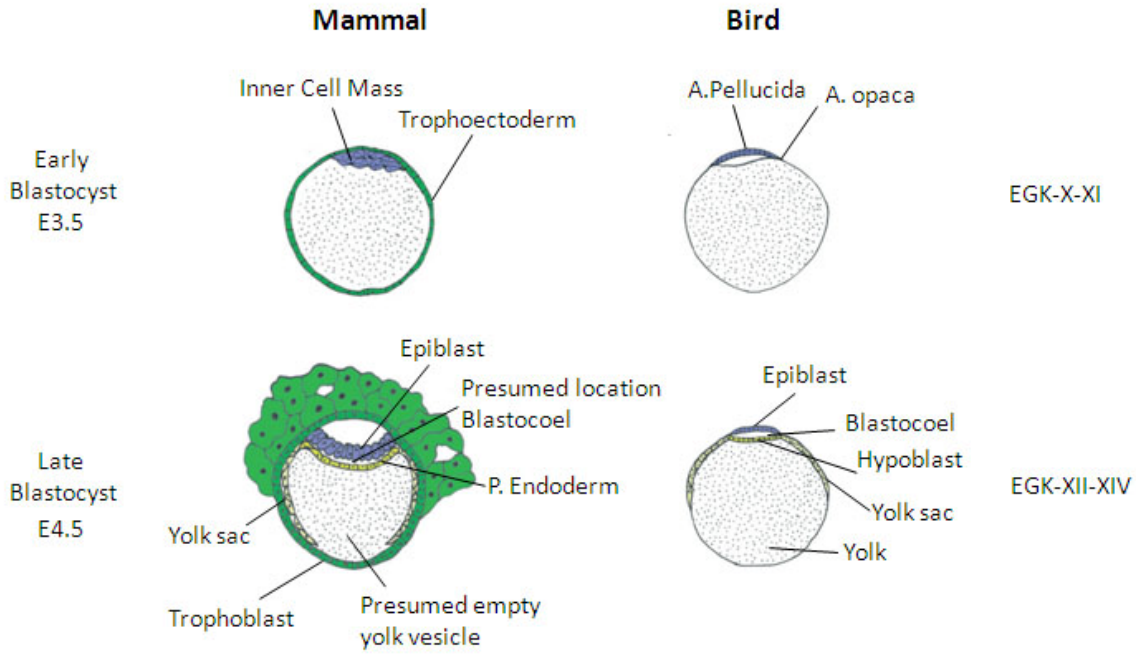


Figure 13. Alignment of early embryonic stages between mammals and birds. In mammals, cells of the epiblast in the late blastocyst (blue) will begin gastrulation into the space between the epiblast and the primitive endoderm (yellow); in birds, at EGK-XII cells from the epiblast (blue) will likewise ingress into the space between the epiblast and the hypoblast (yellow). The cavity in which birds gastrulate can therefore be considered equivalent to the cavity in which mammals gastrulate. In birds this cavity is called the blastocoele, whereas in mammals the blastocoele is the cavity beneath the primitive endoderm, this cavity should be viewed as an empty yolk vesicle instead. Thus the traditional (and incorrect) nomenclature came about from the incorrect alignment of the early mammalian blastocyst (top left) with the chick EGK-XII-XIV (bottom right). Additional parallels can be found between mammals and birds in the formation of the yolk sac from primitive endoderm and hypoblast cells (yellow), respectively. *Modified from* (O'Farrell *et al.*, 2004).

The chick is thus an ideal model for comparing with mouse: not only can we make genomic comparisons and conduct functional assays; we can also make meaningful comparative expression analyses. Such a comparison can thus reveal whether the genetic control of embryonic pluripotency is a conserved feature of mammals and avians or, on the contrary, is a mammalian novelty.

OBJECTIVES

1. Explore the evolutionary history of *Oct4* in vertebrates

Before studying the conservation between mouse and chick embryonic pluripotency networks (objectives 2, 3, 4, 5), we aimed to clarify by phylogenetic, synteny and orthology analysis, the relationships of *Oct4* with its homologues in other vertebrates.

2. Compare the expression of the chick orthologues of mouse EP-GRN core factors and downstream targets

We set out to analyse these genes by in situ hybridization at chick pre-gastrulation stages. This analysis was then extended by comparing the global expression profiles obtained by microarrays of early chick embryos and their derived cells with published data on the mouse pluripotent state.

3. Search for the presence of chick orthologues of other mouse genes downstream of the core factors or of genes considered markers of pluripotency

We searched for these chick orthologues by means of synteny analysis.

4. Analyse the conservation between mouse and chick of the regulatory targets of the core EP transcription factors

We aimed to study, by multi-species genomic alignments, the overall evolutionary conservation of non-coding regions bound in mouse ES cells by the core EP factors. To refine our analysis, we investigated the conservation of selected core EP factor binding sites in functional assays in ES cells and blastocysts.

5. Analyse the conservation of *Nanog* function and its network during evolution

After studying the overall conservation of the embryonic pluripotency network, we focused on one of the core factors, *Nanog*. We aimed to study the functional equivalence of *Nanog* orthologues in ES cells and to compare Nanog interactions between mouse and chick by overexpressing this gene in early embryos.

MATERIALS & METHODS

1. Animals models

1.1 Chick

Fertilized eggs were obtained from poultry farm *Gibert*, Tarragona, and incubated at 38°C 100% humidity for a variable period in relation to the stage of interest. Pre-gastrulation embryos were staged according to the Eyal-Giladi and Kochav (EGK series) (Eyal-Giladi and Kochav, 1976) and post-gastrulation were staged according to the Hamburger and Hamilton (HH series) (Hamburger and Hamilton, 1992).

1.2 Mouse

Double transgenic Nanog overexpressing line (Tg.TetOP-Nanog; ROSA26::rtTA) (Fig 1) was a kind gift from Dr. M. Serrano (*CNIO*, Madrid, Spain) and was generated by Dr K. Hochedlinger (*Harvard University*, Boston, USA) using the same strategy described for transgenic mice that ectopically overexpress Oct4 (Hochedlinger *et al.*, 2005). The M2 reverse tetracycline transactivator requires doxycycline for binding to the tetracycline operator (TetOP) consequently activating a CMV minimal promoter and thus, the over-expression of Nanog (Fig 1).

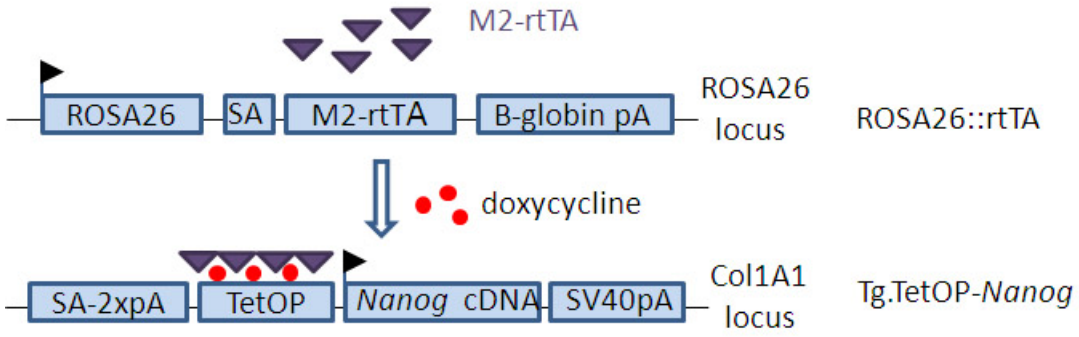


Figure 1. Schematic representation of transgenes used to produce *Nanog*-inducible mice. The M2 reverse tetracycline transactivator (M2-rtTA) was targeted to the ROSA26 locus for constitutive expression (M2-rtTA) under the control of ROSA26 promoter. A cassette containing *Nanog* cDNA and the tetracycline/doxycycline-responsive element (TetOP) with a CMV minimal promoter was targeted downstream of the collagen type 1 alpha 1 locus (Col1A1) by frt/Flpase-mediated site-specific integration. SA, splice acceptor; pA, polyadenylation signal. Black flags indicate transcriptional start sites. *Modified from* (Hochedlinger *et al.*, 2005).

Nanog overexpression was induced by adding 0.2 mg/ml doxycycline in the drinking water given to pregnant females during the three days immediately before embryo extraction. Mouse embryos were staged according to Kaufmann (Kaufmann, 1992).

For genotyping, the Tg.TetOP-*Nanog* transgenic allele was detected with a specific primer, col/frtC1 and a common primer, col/frt-B. The wild type (WT) allele was detected with a specific primer, col/frtA1 and the common primer. The PCR product for the transgenic allele was 551bp in length while for the WT allele was only 331bp. The ROSA26::rtTA transgenic allele was detected with a specific primer, B and a common primer, A. WT allele (+) was detected with specific primer, C and the common primer. The PCR product for the transgenic allele was 300bp in length while for the WT was 500bp (Table 1).

Transgene	Primer name	Sequence	Allele detected
Tg.TetOP-*Nanog*	col/frt-B	5'CCCTCCATGTGTGACCAAGG 3´	WT, Transgenic
	col/frtA1	5'GCA CAGCAT TGC GGA CATGC 3´	WT
	col/frtC1	5'GCAGAAGCGCGGCCGTCTGG 3´	Transgenic
ROSA26::rtTA	A	5'AAAGTCGCTCTGAGTTGTTAT 3´	WT, Transgenic
	B	5'GCGAAGAGTTTGTCCTCAACC 3´	Transgenic
	C	5'GGAGCGGGAGAAATGGATATG 3´	WT

Table 1. Primers used for *Nanog* inducible mice line genotyping. Tg.TetOP-*Nanog*, tetracycline operator and *Nanog* cDNA inserted into the 3' UTR of the *Col1a1* locus; ROSA26::rtTA, reverse tetracycline transactivator (M2-rtTA) downstream of the ROSA26 promoter.

The transgenic line was maintained as double heterozygous breeding pairs (Tg.TetOP-*Nanog* / +) (ROSA26::rtTA/+). Heterozygous males were mated with heterozygous or wild type females. The breed could be Nanog overexpressing embryos (named Nanog)

which had both transgenic alleles (Tg.TetOP-*Nanog* /+) (ROSA26::rtTA/+) or non overexpressing embryos (named controls) with three possible genotypes: (+/+) (ROSA26::rtTA/+), (Tg.TetOP-*Nanog* /+) (+/+) or (+/+), (+/+). We grouped these three control genotypes as we did not observe differences between them.

2. In situ hybridization

Whole mount in situ hybridizations of chick and mouse embryos were carried out as described (Acloque *et al.*, 2008; Ariza-McNaughton and Krumlauf, 2002). For early chick embryos (EGK-X), an average of 10 embryos was processed in parallel with later stages. This was important to confirm the reported gene expression patterns at later stages and to ensure that early embryos were developed for the appropriate time. Sense probes were used as negative controls. For chick and mouse Nanog overexpressing embryos, controls were processed in parallel for the same amount of time. Sources of probes used are described in Table 2. After in situ hybridization, selected chick embryos were embedded in paraffin and 0.7 micron sections were cut on a microtome.

3. Inmunohistochemistry

A monoclonal antibody raised in rabbit against the chick Vasa homolog protein (Tsunekawa *et al.*, 2000) was kindly provided by Dr. Toshiaki Noce (*Mitsubishi-Kasei Institute*, Tokyo, Japan). After fixing in 4% paraformaldehyde in PBS and dehydrating in 100% methanol, chick embryos were rehydrated in PBS containing 0.1% Tween-20. Embryos were then blocked for one hour in blocking solution (50mg blocking reagent (Roche), 100mM Maleic Acid, 150mM NaCl, 400ml water adjusted with NaOH at ph 7.5) diluted 1:100 in B1 solution (100mM TrisCl ph 7.5, 150mM NaCl, 0.1% Triton X 100). Primary antibodies were added in fresh 1% blocking solution at the following diultions: anti-Vasa, 1000 fold; mouse anti-GFP (JL-8 Clontech), 125 fold dilution. The embryos were incubated in this solution overnight at 4°C. After six 1 hour washes with B1 solution secondary antibodies were added and the embryos were incubated overnight at 4°C. Secondary antibodies were anti-rabbit Alexa 568 (A11011, Invitrogen) and anti-mouse Alexa 488 (A11029, Invitrogen). After five 1 hour washes in B1 solution, embryos were examined under the fluorescence microscope.

Gene Name	Species	Source
Oct4	*M.musculus*	Dr. T. Rodríguez (*MRC*, London, UK)
Pou2-r	*G.gallus*	PCR, Primers: F-Pou2, CATGTGCAAGCTGAAGCCACTGCT R-Pou2, TCAGTGGCTGCTGTTGTTCATGGAG
Fgf4	*G.gallus*	Dr. P. Bovolenta (*Instituto Cajal*, Madrid, Spain)
Nodal	*G.gallus*	Dr. P. Bovolenta (Instituto Cajal, Madrid)
Sox2	*M.musculus*	Dr. T. Rodríguez (*MRC*, London, UK)
Sox2	*G.gallus*	MCR Geneservices (Boardman *et al.*, 2002), clone ChEST878b12, acc. nº BU282995,
FoxD3	*G.gallus*	MCR Geneservices (Boardman *et al.*, 2002), clone ChEST68l3, acc. nº BU128393,
Phc1	*G.gallus*	MCR Geneservices (Boardman *et al.*, 2002), clone ChEST49d22, acc. nº BU219008,
Sox3	*G.gallus*	Dr. H. Kondoh (*Graduate school of Frontier biosciences*, Osaka, Japan)
Sox1	*G.gallus*	Dr. H. Kondoh (*Graduate school of Frontier biosciences*, Osaka, Japan)
Eomes	*M.musculus*	Dr. T. Rodríguez (*MRC*, London, UK)
Eomes	*G.gallus*	(Pernaute *et al.*, 2010)
Cdx2	*M.musculus*	Dr. T. Rodríguez (*MRC*, London, UK)
Cdx2	*G.gallus*	(Pernaute *et al.*, 2010)
Fgf8	*M.musculus*	Dr. J.J Sanz-Ezquerro (*CNIC*, Madrid, Spain)
Fgf8	*G.gallus*	Dr. M. Ros (*IBBTEC*, Santander, Spain)

Table 2. Probes used for mouse (*Mus musculus*) and chick (*Gallus gallus*) in situ hybridizations

4. Microarrays

4.1 Sample preparation

For chick microarrays 20 HH6 or 30 EGK-X stage embryos were dissected for each of three independent replicates. Blastoderm-derived cells (bdC) were isolated as described (Petitte *et al.*, 2004). Briefly, 20 intact stage EGK-X area pellucidas for each of three independent replicates were placed into tubes containing PBS and 2% chick serum. After centrifugation PBS was replaced with trypsin 0.05% and incubated for 5 min at room temperature. Dispersed cells were plated into 35 mm tissue culture dishes and cultured for 48 hr in DMEM containing 10% FBS, 2% chick serum, 100U/ml penicillin, 100ug/ml streptomycin and LIF. The cultured cells were positive for alkaline-phosphatase (Fig 2) and showed high levels of expression, as determined by qPCR, of *Pou2-r* and *Nanog* as previously described (Lavial *et al.*, 2007; Pain *et al.*, 1996) but undetectable levels of *Sox2* in agreement with our whole mount in situ analysis (not shown). RNA was extracted from the cells using the RNeasy kit (Qiagen). Integrity and concentration of purified total RNA was determined by a 2100 Bioanalyzer (Agilent) and Nanodrop measurement (Agilent, ND-1000 spectrophotometer).

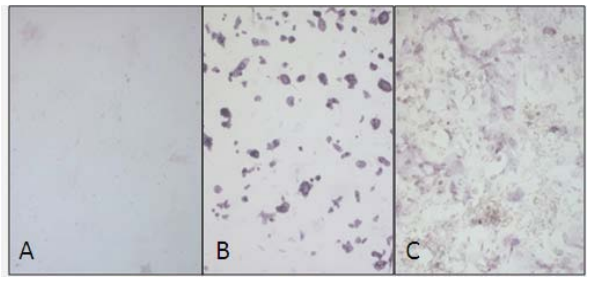


Figure 2. Alkaline phosphastase staining in 2-day cultures of (A) mouse fibroblasts (negative control), (B) mouse ES cell colonies (positive control) and (C) chick blastoderm-derived cells.

For mouse microarrays, stage 9.5 mouse embryos were extracted and stored individually at -80°C until processed. Embryos were genotyped using the yolk sac, which was set aside during extraction. RNA was extracted from a pool of three Nanog overexpressing (Nanog) or non overexpressing embryos (controls) using the RNeasy kit (Qiagen). In total we obtained four replicates (of 3 embryos each) for each condition. Half of the replicates were pools of female embryos and the other half males, in case further analysis related with the sex was required.

4.2 Microarray analysis

For chick microarrays, RNA was amplified and hybridized to the chick (V2) 4x44K Gene Expression Microarray (G2519F, Agilent Technologies) and for mouse, to the mouse whole genome 4x44K Gene Expression Microarray (G4122F, Agilent Technologies); scanning and image analysis was performed following the manufacturer's instructions. Preprocessing and statistical analysis of data was carried out at the CNIC Genomics Unit following standard procedures (Bolstad *et al.*, 2003; Smyth, 2004).

4.3 Data processing

For chick microarrays three way comparisons of the expression data were performed with cut-offs of corrected p-value < 0.01 and -2>logFC>2 (log fold change) to identify those genes overrepresented in one sample versus the other. Sets of mouse genes that respond to downregulation of Oct4, Sox2 or Nanog by RNAi or overexpression of Nanog in ES cells were obtained from Sharov et al (Sharov *et al.*, 2008). Genes downregulated in response to RNAi or upregulated in the case of Nanog overexpression were considered as activated by the core pluripotency factors, and as such a signature of the embryonic pluripotent stage. Genes upregulated in response to RNAi or downregulated in the case of Nanog overexpression were considered as repressed by the core pluripotency factors, and as such a set of genes that should not be overexpressed in the embryonic pluripotent stage. The percentage of the orthologues of over-represented chick genes in each sample (HH6, EGK-X, bdC) present in each mouse set was calculated independently for each EP factor and jointly for all factors, obtaining similar results (Table 3). The sets of genes involved in different aspects of pluripotency were obtained from (Tang *et al.*, 2010) and analysed as described above (Table 4).

	Oct4		Sox2		Nanog		Core EP-TF		Total
	up	down	up	down	up	down	up	down	
HH6	213	156	131	234	76	155	303	354	1853
EGK-X	239	189	153	240	71	156	337	376	2089
bdC	153	238	87	303	84	192	245	438	1826
Total	3617	2580	2035	3277	902	1802	4991	5314	

Table 3. Genes overexpressed in chick HH6 and EGK-X embryos, and blastoderm-derived cells (bdC), whose mouse orthologues are up or downregulated in response to interfering with the endogenous expression of the core EP transcription factors Oct4, Sox2 and Nanog (Sharov *et al.*, 2008).

	HH6	EGK-X	bdC	Total
pluripotent genes	99	100	77	1408
ES specific	317	249	197	2488
ICM specific	236	221	231	2405
represors of pluripotency	57	61	89	703
self-renewal	27	16	12	229
maintainence pluriotency	36	44	28	642
Total	1853	2089	1826	

Table 4. Genes overexpressed in chick HH6 and EGK-X embryos, and blastoderm-derived cells (bdC), whose mouse orthologues are included in sets of genes defined in relation to embryonic pluripotency (Tang *et al.*, 2010).

For mouse microarrays, a corrected p-value cut-off of 0.05 was set to identify genes whose expression was significantly differentially regulated in the Nanog overexpressing mouse embryos. Based on the log FC, genes were sorted into Nanog upregulated (positive logFC) or Nanog downregulated (negative logFC). The top networks and canonical pathways were generated by Ingenuity pathway analysis (Ingenuity® Systems, www.ingenuity.com). A dataset containing gene identifiers and corresponding expression values was uploaded into the application. Each identifier was mapped to its corresponding object in Ingenuity's Knowledge Base. Network Eligible molecules (those with corrected p-value cutoff of 0.05), were overlaid onto a global molecular network developed from information contained in Ingenuity's Knowledge Base. Networks of Network Eligible Molecules were then algorithmically generated based on their connectivity (Results Fig. 27).

5. Sequence analysis

Chromosomal positions and genomic organization of chick and other vertebrate genes and genomic sequences from the *Fgf4* and *Sox2* locus were obtained from the latest release of the Ensembl genome browser (www.ensembl.org).

Synteny analysis was aided by the use of the Genomicus browser, developed by the Dyogen Lab in Paris (www.dyogen.ens.fr) (Results Fig. 2, 10). Orthology maps were constructed by examining the presence of 1-to-1 orthologues in different species (opossum, chick, *Xenopus* and zebrafish) of all genes in a 10 Mb region surrounding mouse *Oct4* and *Nanog*. Mouse genes with no orthologues in the species examined were not used, and when multiple matches to mouse genes were found in other species these were reduced to only one (Results Fig. 3). For phylogenetic analysis, sequences were aligned using Clustal X (Thompson *et al.*, 1997) and edited using the BioEdit tool (Hall, 1999). The alignment was used to construct an unrooted neighbour-joining bootstrapped tree (Felsenstein, 1996; Saitou and Nei, 1987) and drawn with Treeview (Page, 1996) (Results Fig. 4).

Multi-species genomic alignments were performed using Vista tools (Frazer *et al.*, 2004). Sequences surrounding the Sox2/Oct4 sites and the multi-specie Nanog predicted proteins were aligned using Clustal X (Thompson *et al.*, 1997), and edited with BioEdit (Hall, 1999).

Genomic regions bound by Oct4 and Nanog in mouse ES cells were obtained from the ChIPseq data of (Marson *et al.*, 2008). We included in the analysis regions bound together by four (Oct4, Nanog, Sox2 and Tcf3; 814 regions), three (Oct4, Nanog and Sox2; 638 regions) or two (Nanog and Oct4; 236 regions) core EP-GRN transcription factors, located at 8 kb or less from known genes. We obtained similar results if each of these subsets were analysed independently. Genomic regions bound by Gli3 in the mouse limb were obtained from the ChIP-seq data from (Vokes *et al.*, 2008). For our analysis we used the 200 top-scoring regions as described in that paper. The coordinates for both data sets were converted from the mouse mm8 (February 2006) to the mm7 (August 2005) assembly and visualized using the Vista Tracks in the UCSC Browser (Frazer *et al.*, 2004). These data sets were compared with rat, human, dog and chick. Each genomic region was individually analysed: a region was scored as

conserved if it contained at least one >100 bp segment showing over 70% similarity in intergenic, intronic or non-translated regions. Conserved regions overlapping coding exons were not included. Regions were scored as mammalian specific if they were conserved between mouse and at least one of the two non rodent species analysed (human and dog) (Results Fig 11A, B). A list of Ensemble transcript identifiers of those genes mapping in the vicinity (8kb or less) of the Oct4-Nanog bound regions (Marson *et al.*, 2008) was used to extract from BioMart (www.biomart.org) the number of chick orthologues in the Oct4-Nanog data set (Results Fig. 11C).

6. Constructs

6.1 Reporter constructs

Genomic fragments corresponding to the enhancers from mouse *Sox2* and *Fgf4* and the equivalent chick regions were amplified by PCR using the following BAC templates (http://bacpac.chori.org/) and primers: mouse *Sox2*, BAC RP24-140C7, AAGGCACCAAGAACCAGAAAT and TCAGCAAGTCCTCTCTGGGTA; mouse *Fgf4*, BAC RP23-294B14, GGTGAAAATATGCACGACCAG and TGAATGCTTCTCTTTGGATGG; chick *Sox2*, BACs CH261-178A15 and CH261-110M, AGATTCAGGCATTTGATCTCG and AAACAAGCGGTGAATTCCTCT. Chick *Fgf4* was amplified from genomic DNA with ATGGTGTGAAAAGTGGCAAAG and TTTGGTGCAGTATTTGGAAGG. The fragments were cloned in pGem-T Easy vector (Promega) and mutated versions were generated by site-directed mutagenesis (Mutagenex Inc., NJ USA) (Table 5).

Name	Wild type	Mutated version	Changes
Mouse-to-chick *Fgf4*	CTTTGTTTGGATGCTAAT	aTTTGTTctatTGg-AAT	6 substitutions and 1 deletion
Chick-to-mouse *Fgf4*	ATTTGTTCTATTGG-AAT	cTTTGTTtggaTGctAAT	6 substitutions and 1 insertion
Mouse-to-chick *Sox2*	CATTGTGATGCATAT	tggTGTaAaGaAacc	9 substitutions
Chick-to-mouse *Sox2*	TGGTGTAAAGAAACC	catTGTgAtGcAtat	9 substitutions

Table 5. Mutated versions of mouse and chick Fgf4 and Sox2 enhancers. The changes introduced are indicated in red and lower case.

Wild type and mutated enhancer fragments were then cloned into a vector containing a human minimal beta-globin promoter upstream of a *lacZ* reporter linked to a SV40 polyadenylation signal (Yee and Rigby, 1993) (This vector was a kind gift of Dr. R. Krumlauf; *Stowers Institute*, Kansas, USA.). This minimal beta-globin promoter is barely active, and needs of an enhancer, cloned upstream or downstream, to drive the expression of the *lacZ* reporter. Thus reporter activity provides a readout of the regulatory potential of the enhancer fragment.

These constructs were used for both ES cell transfections and, once linearized and the plasmid backbone removed, for the generation of transgenic embryos. As a positive control we used the previously described *Oct4* distal enhancer element (*Oct4*-DE) (Yeom *et al.*, 1996) that was cloned following the same strategy, using BAC RP23-152G18 as template and primers AGCGGCCGCCTCTGCTACATGTAAATTTGTCT and AGCGGCCGCCTAAACAAGTACTCAACCCTTGAA.

6.2 Expression constructs

A partial chick *Nanog* cDNA had been previously cloned in our group and used for in situ hybridizations (Canon *et al.*, 2006). For further analysis in ES cells and for

electroporation of chick embryos we needed the full-length sequence. We therefore obtained cDNA from total RNA from whole chick embryos at stage HH20 (Hamburger and Hamilton, 1992). Primers were designed using the predicted sequence (Table 6). For PCR a high fidelity polymerase, Platinum Taq (Invitrogen), was used and the amplification product cloned first into pGEM-T easy vector (Promega) and later into pPyCAGIP vector for stable transfections. This vector, a kind gift of Dr. I. Chambers (*CMR*, Edinburgh, UK), contains a CAG promoter (a combination of chick beta-actin promoter and cytomegalovirus immediate-early enhancer) followed by a stuffer. The cDNA is cloned in place of the stuffer fragment within a transcription unit linked to a puromycin resistance gene through IRES. The puromycin resistance gene allows elimination of untransfected ES cells (Chambers *et al.*, 2003).

Opossum *Nanog* was predicted in the opossum *(Monodelphis domestica)* genome annotation at reference ID ENSMODG00000017979 on chromosome 8, Ensembl vs49 2008. To obtain the full-length opossum *Nanog* cDNA, we reverse transcribed RNA from stage Mc (McCrady) 29 genital ridges (Mate *et al.*, 1994), a kind gift from Dr. K. Smith (*Duke University*, Durham, USA). A GeneRacer ^{TM} technique (Invitrogen) was used to amplify the 5’ and 3´ ends with primers designed from ``in silico´´ prediction using GenScan00000034861 (Table 6). Subsequently, primers were designed from the sequence obtained by GeneRacer ^{TM} and used to isolate full length *Nanog* sequence (Table 6). All PCRs were performed using Platinum Taq polymerase (Invitrogen) and the full length sequence was cloned initially into pGEM-T easy (Promega) and later into pPyCAGIP. Mouse *Nanog* cDNA was a kind gift of Dr I. Chambers (*CRM*, Edinburgh,UK) and was also cloned into pPyCAGIP.

For	Primer name	Sequence
Full-length *Gg- Nanog*	F-fullGg	5´GCTCACCGCCAGCAACGGC 3´
	R-fullGg	5´GGATCTATGAGTCAAGACCCTAG 3´
5´ end *Md-Nanog*	F-GSPMd	5´CTGGTTGCTCCATACTGGAAGAGT 3´
	R-GSPnestedMd	5´CGGATCTGCTGTGGGCTTAGGTAT 3´
3´ end *Md-Nanog*	F-GSP-Md	5´CCTGATTCTGCCACTAGCCCTACT 3´
	R-GSPnestedMd	5´CTTCCAGTATGGAGCAACCAGACT 3´
Full-length *Md- Nanog*	F-fullMd	5´CCTTCCAGCATCTCCTCTAAAAC 3´
	R-fullMd	5´GATAAGGAGCCCTGGAAAGAGTA 3´

Table 6. Primers used for cloning chick and opossum *Nanog*. F, forward and R, reverse.

To analyse the pluripotency enhancers *Fgf4* and *Sox2*, mouse *Oct4* and chick *Pou2-r* cDNAs were cloned into pCAGGS vector. These were kind gifts from Drs. H. Kondoh (*Graduate school of frontier biosciences*, Osaka, Japan) and J. Rodríguez-León, (*CMRB*, Barcelona, Spain) respectively. The pCAGGS vector contains a ubiquitous and strong chick beta-actin promoter that drives cDNA expression (Niwa *et al.*, 1991).

For electroporation, the chick full-length *Nanog* cDNA was cloned from pGEM-T easy vector (Promega) into pCAGGS/SE, a kind gift from Dr. M Torres (*CNIC*, Madrid, Spain). The pCAGGS/SE vector includes a greater number of restriction sites 3´ of the promoter than the original pCAGGS vector (Stuhmer *et al.*, 2002); the *Nanog* cDNA was cloned into the EcoRI site.

7. ES cell culture and transfections

Mouse ES cell line E14 was used to test the activity of the *Fgf4* and *Sox2* EP enhancers. These ES cells were grown on gelatine-coated plates without feeders in Dulbecco´s modified Eagle's medium supplemented with LIF, 15% fetal calf serum, 1mM glutamine, 1mM sodium pyruvate, non essential amino acids and 100µM beta-mercaptoethanol. Transient transfections were performed as previously described (Robertson *et al.*, 2006) using in this case 12 µl of Lipofectamine 2000 (Invitrogen) and

2.5µg of each construct. The pPyCAG-GFP vector (a kind gift from Dr I. Chambers, *CMR* Edinburgh, UK) was co-transfected in all cases as an internal control for efficiency. Cells were examined 24 h later and for each well, cells expressing *lacZ* and GFP were counted in four random and independent fields (average 300 cells per field). The number of *lacZ* positive cells was normalized with respect to GFP positive cells, and the value obtained for the *Oct4* distal enhancer was assigned the value of 1. Negative controls included a mock transfection (only pPyCAG-GFP) and the empty vector. Three independent transfections were carried out for each construct.

To test the function of mouse *Oct4* and chick *Pou2-r* on the *Fgf4* or *Sox2* pluripotency enhancers, we used mouse E14 and ZHBTc4 ES cells (a kind gift from Dr I. Chambers, *CMR*, Edinburgh, UK). E14 ES cells were transfected under the same conditions as above with different combinations of the overexpressing constructs: mouse *Oct4* or chick *Pou2-r* cloned into pCAGGS, together with the different versions of the *Fgf4* and *Sox2* enhancers. Relative activity of enhancer elements was determined as the proportion of *lacZ* positive cells for each combination of reporter and expression constructs. ZHBTc4 ES cells, in which *Oct4* expression can be repressed by the addition of tetracycline (Niwa *et al.*, 2000), were cultured as above and transfected under the same conditions with the *Oct4*-DE construct using mRFP as reporter, together with mouse *Oct4*, chick *Pou2-r*, or empty pCAGGS vector. When required, tetracycline (10ng/ml) was added after 24 hours, and reporter activity was measured 48 hours after transfection.

The *Nanog*-null mouse ES line RCNβH(t) (Chambers *et al.*, 2007), a kind gift from Dr. I Chambers (*CRM*, Edinburgh, UK), was used to analyse the functional equivalence between mouse, opossum and chick Nanog proteins. The *Nanog*-null ES cells were grown under the same conditions as the E14 cells but with the addition of hygromycin B (50µg/ul) to select the cells in order to maintain the mutation. For stable transfections 10^7 cells were electroporated at 3uF, 0.8 kV (gene pulser, Biorad) with 100ug of vector linearized with ScaI. As controls, we included empty pPyCAG and mock transfection (only electroporation). Cells were plated at 5 x 10^5 cells per 9cm plate. Selection started after 48 hours with the addition of 1ug/ml puromycin to culture medium with or without LIF. Puromycin was administered every two days until mock

plates were clear, usually in 12 days. Colonies were stained with an Alkaline phosphatase kit (Sigma) and classified as fully differentiated, mixed or undifferentiated.

8. Embryo transgenesis

Transient transgenic blastocysts were generated by pronuclear injection as described (Nagy *et al.*, 2003). Each construct was micro-injected at a concentration of 4ng/µl and embryos were cultured in microdrops of M16 medium (Sigma) covered with mineral oil (Sigma) at 37°C, 5% CO2 until blastocyst stage, fixed in 1% formaldehyde, 0.2% glutaraldehyde, 2mM MgCl2, 5mM EGTA, and 0.02% Igepal for 5 min at room temperature, washed in PBS, and stained for betagalactosidase activity as described (Andras Nagy 2003) for 24 hours at room temperature in the dark. A minimum of 50 blastocysts were used per construct to calculate the percentage of positive embryos. The empty vector containing only the minimal promoter and the *lacZ* reporter was used as a negative control, as positive control we used the *Oct4-DE*.

9. Electroporation

Chick embryos were electroporated with *Nanog* in pCAGGS/SE at concentrations ranging from 0.1 to 2ug/ul together with pCAGGS-GFP (0.2ug/ul) to monitor the area and efficiency of electroporation. Control embryos were electroporated with empty vector and pCAGGS-GFP. Electroporation was performed as described (Uchikawa *et al.*, 2004). Briefly the egg was broken into a dish, and the HH4 stage embryo was adhered to a ring of filter paper through the vitelline membrane. The embryos were placed upside down onto the electrode (cathode) and the DNA solution injected between the blastoderm and the vitelline membrane using a glass capillary. An anodal electrode was quickly placed on the hypoblast side of the embryo and electroporation performed with 5 pulses of 7 V for a duration of 50msec and with intervals of 100msec (electroporator Intracel TSS20). The embryos were then incubated on agar culture medium made of liquid albumen, 0.15% glucose, 61.5mM NaCl and 0.3% noble agar at 38°C and 100% humidity. GFP expression can be detected about 4 hours after electroporation and for as long as 72 hours afterwards. The embryos of interest were

then fixed in 4% paraformaldehyde and stored at –20°C until processing for in situ hybridization or inmunohistochemistry (Fig 4).

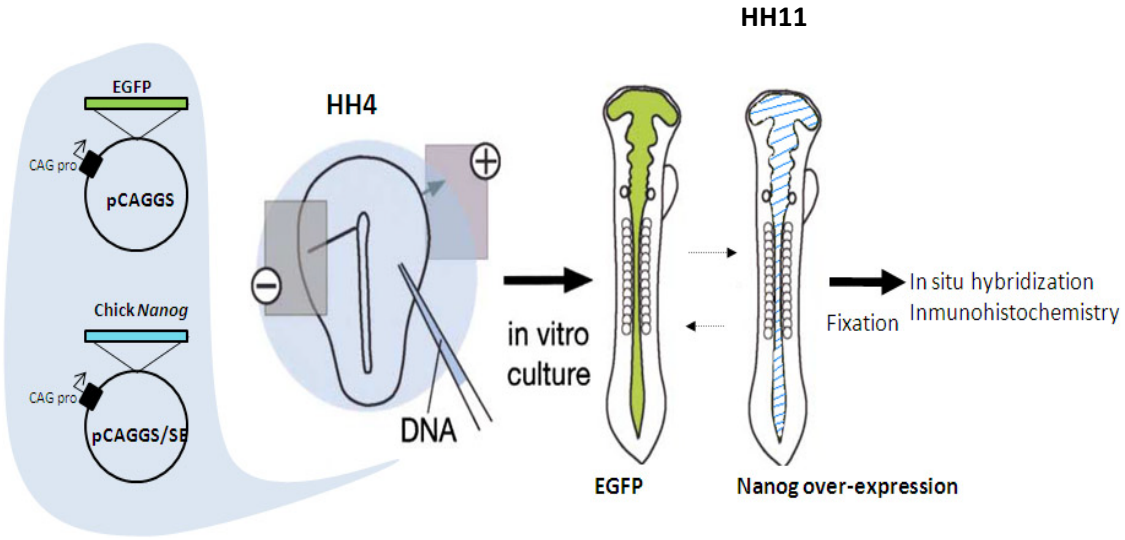


Figure 4. Electroporation of chick embryos to overexpress chick Nanog. Electroporation involves application of an electric field to the embryo that transiently disrupts the stability of the cell plasma membrane, thus creating reversible pores through which the plasmids can be transported into the cytosol .*Modified from* (Uchikawa, 2008).

Nanog overexpression assays in ES cells were performed during a short-term stay at the laboratory of Dr. Ian Chambers, *Center for Regenerative Medicine*, Edinburgh, UK.

Initial chick embryo electroporations, in situ hybridizations and vasa immunohistochemistry were performed during a short-term stay at the laboratory of Dr. Hisato Kondoh, *Graduate school of frontier biosciences,* Osaka University, Japan.

In situ hybridizations, chick microarrays and ES cell culture were done in collaboration with Dr. Susana Cañon. Reporter constructs, ES cell culture and embryo transgenesis were carried out in collaboration with Dr. Bárbara Pernaute and Teresa Rayón.

RESULTS

1. A COMPARATIVE ANALYSIS OF EMBRYONIC PLURIPOTENCY NETWORKS BETWEEN MOUSE AND CHICK

1.1 Identification of an *Oct4* paralogue in chick

At the start of this research project, homologues of the mammalian core factor *Oct4* (official name *Pou5f1*) had been described from various non-mammalian vertebrates (Bachvarova *et al.*, 2004; Burgess *et al.*, 2002; Morrison and Brickman, 2006). All these studies claimed that such genes were true homologues of mammalian *Oct4* based on sequence similarity, early embryonic expression, conservation of synteny (Burgess *et al.*, 2002; Morrison and Brickman, 2006), or functional conservation in tissue culture assays (Lavial *et al.*, 2007; Morrison and Brickman, 2006). However, their relationships were never assessed from a comparative genomic perspective. Therefore we decided to analyse the evolutionary history of *Oct4* in vertebrates before further analysis of the core network in chick.

By in silico analysis of the chick genome we identified a Pou-domain containing gene, with high similarity to *Oct4* and identical to the *cPouV* gene described by (Lavial *et al.*, 2007). We also analysed the opossum *(Monodelphis domestica*) genome, a marsupial, and surprisingly found two genes highly related to *Oct4*. Phylogenetic analysis revealed that one of the opossum peptides was more related to the chick peptide than to the other opossum sequence, the latter grouping with human OCT4 (Fig. 1). The chick gene, opossum gene and other genes related to them were named as *Pou2*-related (*Pou2-r*) and the second opossum gene and related genes as *Oct4*-related (*Oct4-r*) (Fig. 1). We referred collectively to all *Pou2-r* or *Oct4*-r genes as *Pou5*.

We next examined the genomic location of chick and opossum genes and analysed their conservation of synteny with other vertebrate species (Fig. 2). Chick and opossum *Pou2-r* genes are located in a region of conserved synteny with teleost fish and Xenopus *Pou5* genes. This region also shows extended synteny with the mouse genome but however there is no evidence for genes or pseudogenes related to *Pou2-r* in the expected location flanked by *Npdc* and *Fut7* (Fig. 2). However, when we examined the location of opossum *Oct4* we found that it is in a region of conserved

synteny with the mouse *Oct4* gene and the green anole lizard (*Anolis carolensis*). This region is also present in *Xenopus* but harbouring no *Oct4* related gene and is not conserved in teleost fish.

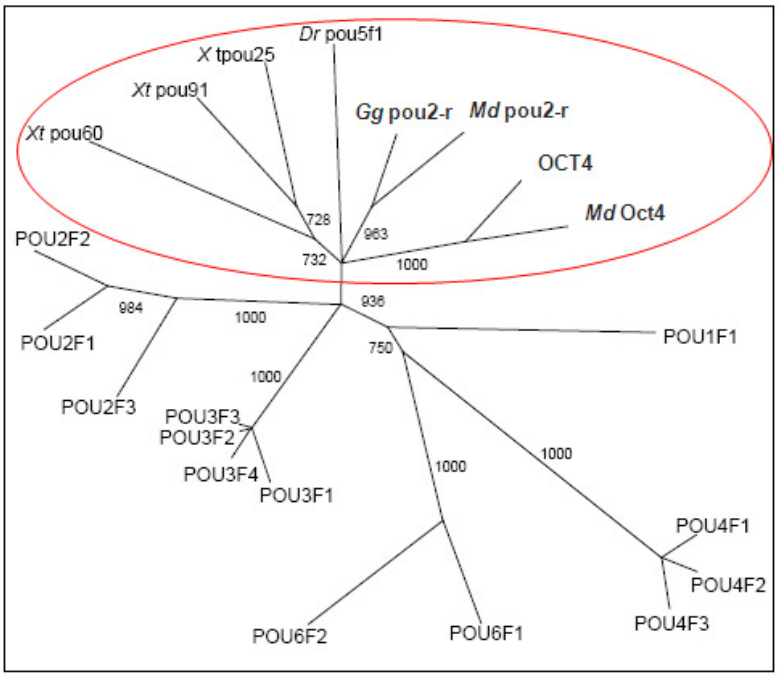


Figure 1. Pou5 phylogenetic tree.

Phylogenetic analysis of the predicted peptide sequences from non-mammalian genes and opossum showed that all grouped together with human OCT4 (red circle) and separated from other human POU factors. Opossum *Md*Pou2-r is more related to chick *Gg*Pou2-r than to the other sequence found in opossum *Md*Oct4. *Xt, Xenopus tropicalis*; *Dr, Danio Rerio*; *Gg, Gallus gallus*; *Md, Monodelphis domestica.*

More surprisingly, searches of trace archives and EST databases of the chick genome detected no orthologues of genes that flank mouse *Oct4* (Fig. 3). Because such genes are present in other vertebrates, it is likely that this region has been specifically lost in the avian lineage. This view is supported by the fact that mammalian *Oct4* is located in the extended MHC Class III region, which is conserved with *Xenopus* but has been lost in chick and other avians (Deakin *et al.*, 2006). Previous claims of conserved synteny between zebrafish and *Xenopus Pou2-r* genes and mammalian *Oct4* (Burgess *et al.*, 2002; Morrison and Brickman, 2006) were surely confused by the fact that there is linkage of both *Pou2-r* and *Oct4* to members of the chloride intercellular channel family (*Clic3* and *Clic1*, respectively; Fig. 2)

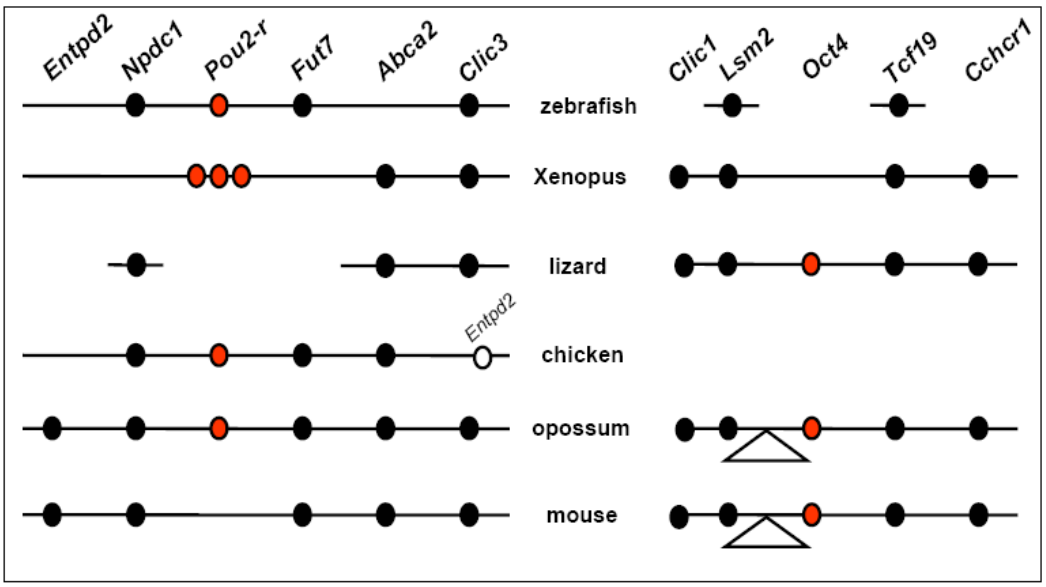


Figure 2. Synteny analysis shows the genomic region containing *Oct4* is lost in avians. Chick *Pou2-r* is located in a region of conserved synteny with that of other *Pou2*-related genes from non-mammalian vertebrates, but different from mouse *Oct4*. The opossum has two *Pou5*-like genes, one in conserved synteny with chick *Pou2-r*, and another with mouse *Oct4*. In the case of the anole lizard, an orthologue of *Oct4* is detected in conserved synteny with mouse, but not an orthologue of *Pou2-r*, but we believe this is due to the lack of genome coverage in the region. Triangles below the opossum and mouse indicate the location of the MHC class III cluster.

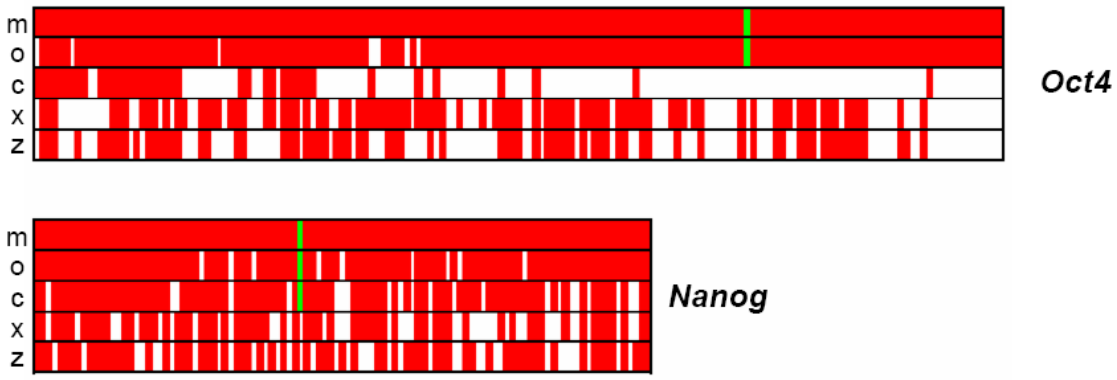


Figure 3. Orthology maps of *Oct4* and *Nanog*. Orthology maps show that the chick genome lacks orthologues for many genes that map in the vicinity of mouse (m) or opossum (o) *Oct4*, but are present in Xenopus (x) or zebrafish (z). In contrast, genes surrounding mouse *Nanog* show a distribution of orthologues in different vertebrates that relates to their evolutionary distance from mouse. The positions of mouse *Oct4* and *Nanog* are indicated by green lines. Presence of orthologues is indicated in red and absence with an empty space.

Such findings mean that the exact time of duplication of the ancestral vertebrate *Pou5* gene cannot be assigned with certainty to the base of the mammalian lineage, since it could have occurred before the divergence of the avian lineage and was subsequently lost as part of the whole region. The existence of the unambiguous *Oct4* orthologue in the lizard genome suggests that the duplication of the ancestral *Pou5* gene occurred before the divergence of amniotes. The exact time of the duplication is unknown; despite our efforts to relate a *Pou5* gene from the amphibian axolotl to *Oct4* or *Pou2*-r genes, we failed due to the lack of genome coverage in the region.

A phylogenetic tree shows the relationships of the species and the duplications and losses during the evolutionary history of *Pou2-r*/*Oct4* (Fig. 4).

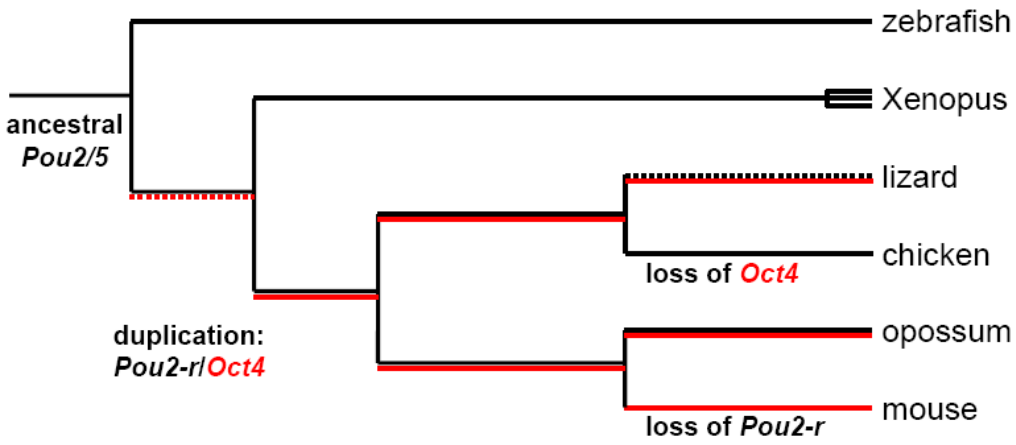


Figure 4. Phylogenetic tree *Pou2-r/Oct4* .The tandem duplication that originated three *Pou2-r* genes in *Xenopus* is indicated. The exact time of the duplication of the ancestral *Pou2/5* gene to give raise to the *Pou2-r* (black line) and *Oct4* (red line) genes is uncertain (indicated as a dashed red line), phylogenetic analysis of the *Oct4*-related gene from the amphibian axolotl places it closer to *Oct4* than to *Pou2-r* genes. However, the lack of genomic data from axolotl precludes completing this analysis with synteny information. In a similar fashion, at this stage we are uncertain about the existence of a lizard *Pou2-r* gene due to lack of genomic information (dashed black line).

We can build an evolutionary scenario in which a duplication of an ancestral *Pou2/5* gene occurred before the divergence of amniotes and subsequent losses occurred in avians (*Oct4*) and eutherian mammals (*Pou2-r*). Consequently, chick *Pou2*-related is not the orthologue of the pluripotency gene *Oct4* and other vertebrate *Pou5* genes, but a paralogue.

1.2 Expression of the core network factors in the early chick

We next decided to examine the expression at pre-gastrulation stages of the chick orthologues of the core EP-GRN. While the expression of *Pou2-r* and *Nanog* had been already described (Canon *et al.*, 2006; Lavial *et al.*, 2007) no studies report the expression in early chick embryos of the other core factor of the network, *Sox2*.

We obtained by in situ hybridization identical results to those previously described from gastrulation to later stages for *Nanog* and *Pou2-r* expression (Canon *et al.*, 2006; Lavial *et al.*, 2007). Both genes are expressed in the epiblast at EGK-X but while *Nanog* expression is confined to scattered cells over the epiblast (the PGCs), *Pou2*-r is expressed throughout the embryo. In contrast, no expression of *Sox2* was detected at the EGK-X until stage HH4 when it is strongly upregulated in the neural plate as reported (Fig. 5) (Rex *et al.*, 1997; Uchikawa *et al.*, 2003).

However *Sox2* belongs to the SoxB1 subfamily of *Sry*-related genes (Schepers *et al.*, 2002) and both other members of this subfamily (*Sox1* and *Sox3*) can partially compensate for *Sox2* in the derivation of mouse iPS cells (Nakagawa *et al.*, 2008), what could point to redundancy in the role of these factors in embryonic pluripotency. We therefore examined the expression of chick orthologues and found that while *Sox1* is not expressed at stage EGK-X, *Sox3* shows low levels of restricted expression in the central epiblast at this stage. Later both are expressed in the neural plate as previously shown (Fig. 6); (Rex *et al.*, 1997; Uchikawa *et al.*, 2003). The low level of expression of *Sox3* mRNA we observed in early chick embryos might not represent a broad expression over the whole of the area pellucida. Published evidence shows that explants from pre-gastrulating chick embryos cultured for 40 hours start to express Sox2 and Sox3 protein only if they were taken from the medial region of the area pellucida, but not from the lateral aspect (Wilson *et al.*, 2001). Therefore, these papers would suggest that already at stage EGK-X, there are differences in the epiblast of the area pellucida in the expression of *Sox2* and *Sox3*, further confirming that *Sox2* (and *Sox3*) is not expressed in the whole epiblast.

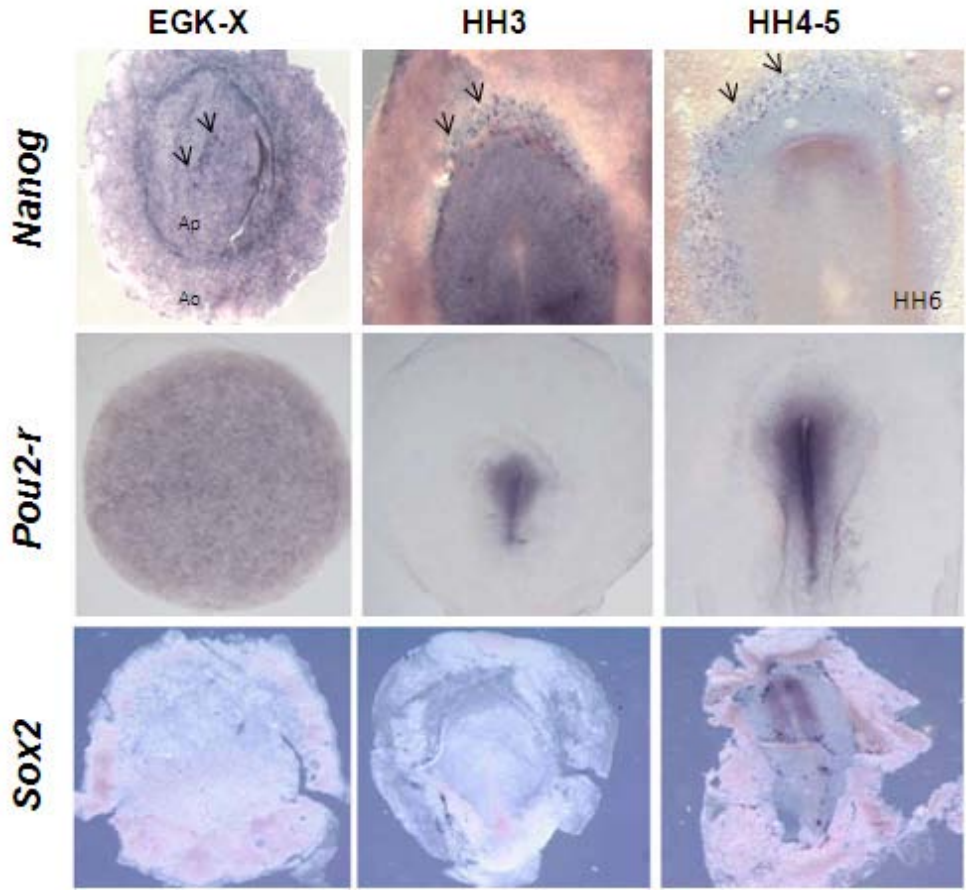
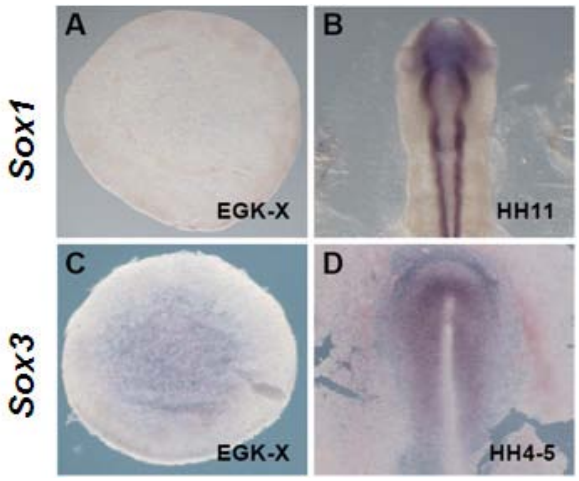


Figure 5. Two of the core factors of the EP-GRN, *Nanog* and *Pou2-r* are expressed in the epiblast in the early chick embryo but no expression is detected for the third, *Sox2*. At stage EGK-X *Nanog* is expressed in scattered cells over the epiblast, the PGCs, later at HH6, expression is confined to the PGCs around the border of the area pellucida and area opaca at the germinal crescent**.** *Pou2-r* is expressed in the earliest stage EGK-X embryos and shortly after is upregulated in the area pellucida. In contrast *Sox2* is first expressed in the neural plate at HH5. Arrows PCGs; Ao, Area Opaca; Ap, Area Pellucida.

Figure 6. Expression of *SoxB1* family members in the chick embryo. *Sox1* is not expressed in the chick embryo at stage EGK-X (**A**), but is clearly expressed at later stages (HH11) in the neural plate (**B**). *Sox3* is expressed at low levels in a restricted fashion in the epiblast at stage EGK-X (**C**) and later can be detected in the neural plate at stage HH4-5 (**D**).



1.3 Expression of other orthologues downstream of the core network in chick

We next studied the expression of orthologues of mouse genes that act in the second layer of the EP-GRN, directly downstream of the core factors, such as the key signaling molecules *Fgf4* and *Nodal* or the transcription factor *FoxD3* which is required for maintaining pluripotency in the epiblast and in ES cells (Hanna *et al.*, 2002). We also analysed the polycomb group gene *Phc1*, because it is a direct target of the core EP factors (Boyer *et al.*, 2005; Loh *et al.*, 2006) , is expressed in ES cells, and maps close to the *Nanog-Gdf3-Dppa3* pluripotency gene cluster (Canon *et al.*, 2006; Clark *et al.*, 2004). Mouse orthologues of all these genes are expressed in the epiblast of the blastocyst (except for *Phc1*, whose expression in the early mouse embryo has not been reported) and in ES cells, but however none is expressed in the pre-gastrulation chick embryo (Stage EGK-X). We detect onset of expression of these genes only at later stages, broadly in line with previously described patterns and domains (Fig. 7) (Chapman *et al.*, 2002; Kos *et al.*, 2001; Shamim and Mason, 1999; Tomotsune *et al.*, 2000).

Orthologues of genes that maintain embryonic pluripotency and early lineages in mouse are not expressed in equivalent territories of the pre-gastrulation chick embryo. These genes therefore might not have the same function in chick as they have in mouse.

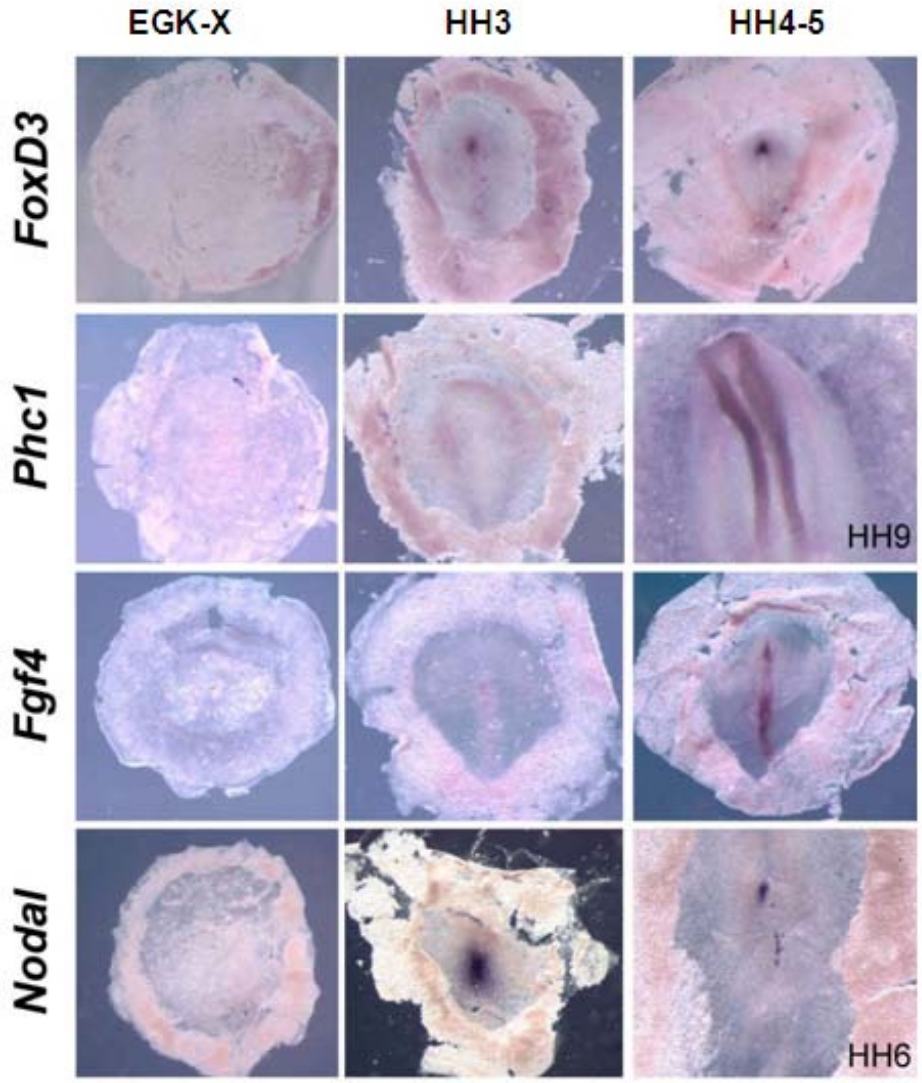


Figure 7. Chick orthologues downstream of the core factors of the mouse EP-GRN genes are not expressed in the epiblast at early pre-gastrulation embryos. *FoxD3* is expressed in the node at HH3 earlier than reported (Kos *et al.*, 2001), *Phc1* in the neural tube at later stages (HH9), a pattern similar to its late expression in the mouse embryo (Tomotsune *et al.*, 2000), *Fgf4* in the primitive streak at HH4 (Shamim and Mason, 1999), and *Nodal* in the early primitive streak at HH3 (Chapman *et al.*, 2002) and later restricted to its anterior portion (HH6).

1.4 Global expression analysis

In order to address in a more global fashion the differences between early mouse and chick embryos we extended our analysis by carrying out a detailed expression profiling of chick blastoderm-derived cells and early developmental stages. We used the pre-gastrulation stage EGK-X and blastoderm-derived cells, equivalent to mouse blastocyst

and ES cells respectively, and representatives of the pluripotency state in mouse. The neural-fold stage HH6 was included as representative of a differentiated state. We compared the profile of chick orthologues of pluripotency-related genes from mouse for each of the samples (HH6, EGK-X, bdC) assuming that if the EP-GRN is conserved between mouse and chick, the EGK-X and blastoderm-derived cell (bdC) sets will be enriched in pluripotency related genes compared to the HH6 set. We carried out the microarrays and compared those genes that were up-regulated in each of the chick samples to those that respond in mouse ES cells to changing the levels of the core pluripotency factors *Oct4*, *Sox2* and *Nanog.* The sets of mouse genes were obtained from {Sharov, 2008 #531} that includes a meta-analysis of various previous papers where such changes in gene expression were analyzed. We found that genes positively regulated by the core factors, and therefore candidates to be part of the EP-GRN, are not over-represented in EGK-X embryos compared to HH6, and even under-represented in blastoderm-derived cells. On the other hand, genes that are negatively regulated by the core factors, possibly involved in differentiation events after the stage of embryonic pluripotency, are not enriched in HH6 embryo compared to EGK-X, and over-represented in blastoderm-derived cells (Fig. 8). We obtained similar results when we calculated the genes positively regulated independently for each EP factor instead for all factors together (Table. 3 Materials and Methods).

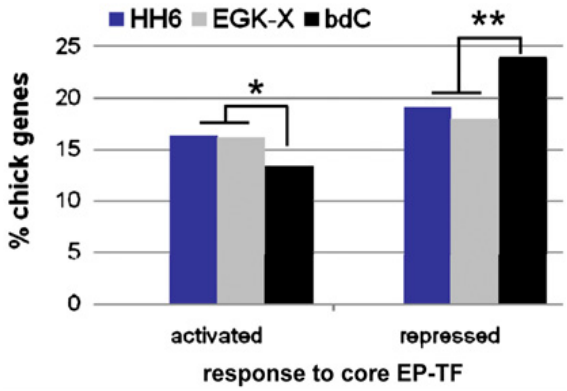


Figura 8. Chick orthologues of mouse EP-GRN genes are not enriched in early pre-gastrulation embryos. Global analysis of gene expression shows that chick orthologues of mouse genes that are activated by core pluripotency factors (EP-TF: Oct4, Sox2 and Nanog) are equally enriched in genes upregulated in HH6 or EGK-X embryos, as compared to blastoderm-derived cells (bdC). Genes repressed by core factors show the same trend only that they are over-represented in upregulated genes in bdC. *, $p < 0.05$; **, $p < 0.01$ (two-tailed Fisher's exact test).

We found similar trends when we compared the chick data set with the results of a recent single-cell analysis of mouse embryonic pluripotent cells (Tang *et al.*, 2010). In this work, the authors compile various data sets of genes that mark different aspects of embryonic pluripotency (pluripotent genes, self-renewal, maintenance of pluripotency, repressors of pluripotency) as well as ES cell and ICM specific genes. Genes defined as pluripotent, ES or ICM specific do not show an over-representation in chick EGK-X samples as compared to the post-gastrulation HH6 embryos. Surprisingly, blastoderm-derived cells show enrichment in genes classified as repressors of pluripotency, as well as a lower proportion of ES-specific or self-renewal genes compared to the HH6 embryo (Fig. 9).

The global expression profiles of pre-gastrulation chick embryo (EGK-X) and blastoderm-derived cells are not more similar to the pluripotent state in mouse embryos than to the profile in the post-gastrulation chick embryo. This strongly suggests the genes that form the mouse EP-GRN are not expressed in a comparable fashion in the early chick embryo.

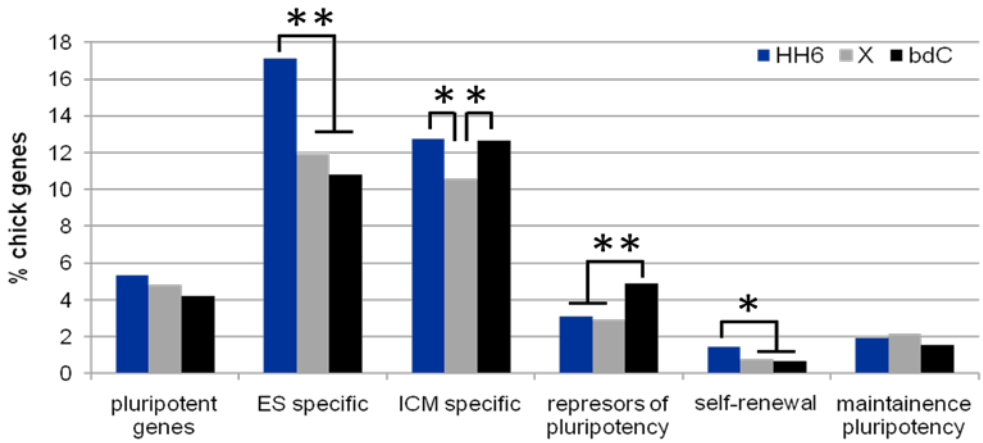


Figure 9. Expression profiling shows no enrichment of orthologues of mouse pluripotency-related genes in the pre-gastrulation chick. The distribution of orthologues of chick genes over-represented in HH6, EGK-X embryos, and blastoderm-derived cells (bdC) in defined sets of pluripotency-related genes (Tang *et al.*, 2010) shows that general pluripotent genes are equally distributed in the three sets of chick genes, while ES and self renewal genes are over-represented in the HH6 set. Genes classified as repressors of pluripotency are over-represented in the bdC set. *, $p < 0.05$; **, $p < 0.01$ (two-tailed Fisher's exact test).

1.5 Several pluripotency-related genes are specific to mammals

We next searched for chick orthologues of mouse genes shown to be downstream of the core factors as part of the EP-GRN, or expressed in similar pattern to these core factors and have been extensively used as markers of the pluripotent state. Among these, we found that in many cases no chick orthologue is present in the available genomic and transcriptional databases. This could be due simply to lack of coverage of genomic and EST sequences and we therefore examined these cases in detail to determine if these genes were unique to the mammalian lineage, as is the case for the pluripotency-associated genes *Gdf3* and *Dppa3*, which map adjacent to *Nanog* (Canon *et al.*, 2006).

The chromosomal regions surrounding the mouse genes *Utf1*, *Tex19.1*, *Dppa2*, *Dppa4* and *Dppa5* conserve synteny with the chick genome, but none of these genes is present in chick, although their immediate neighbours are (Fig. 10). Extensive searches found no evidence for orthologues of these genes in other genomic positions or in other databases. *Utf1* is a chromatin associated factor involved in controlling the initiation of ES cell differentiation and is expressed in the ICM of the blastocyst (Okuda *et al.*, 1998; van den Boom *et al.*, 2007). It is a direct functional target of the Oct4/Sox2 transcriptional complex (Nishimoto *et al.*, 2005), and it has been shown to enhance the efficiency of human iPS cell generation (Zhao *et al.*, 2008). In mouse it is located on chromosome 7, near to genes not involved in pluripotency that are present and conserve synteny with chick, but *Utf1* is specifically missing at that chromosomal position (Fig. 10A). *Tex19.1* and its tandem duplicate *Tex19.2* have been previously described as mammal-specific and expressed in pluripotent stem cells (Kuntz *et al.*, 2008), although this comparison was done with teleost fish and invertebrates and did not include chick. It is noteworthy that these genes are adjacent to another pair of tandem duplicates present in mouse but not in chick, *Sectm1a* and *Sectm1b* (Fig. 10B). These genes belong to the Ig superfamily and are closely related to the neighbouring *Cd7* gene, which is conserved in chick. This indicates that this region has undergone extensive duplication and gene gain during mammalian evolution. *Dppa5*, identified because of having an expression pattern similar to *Oct4*, is flanked by *Omt2b* (*Oocyte*

maturation beta) and *Ooep* (*Oocyte expressed protein homolog*), two other mammals specific genes (Fig. 10C). These genes are expressed in a very similar pattern to *Dppa5* in the embryo and ES cells (Popova and Morris, 2004; Tanaka *et al.*, 2006), and furthermore *Ooep* and *Dppa5* belong to the same family (Pierre *et al.*, 2007). *Dppa2* and *Dppa4* are tandem duplicated genes expressed in the ICM and later in primordial germ cells (Maldonado-Saldivia *et al.*, 2007), and interestingly forced expression of *Dppa4* in ES cells drives them to a primitive ectoderm lineage (Masaki *et al.*, 2007). Furthermore, *Dppa4* is a direct target of the *Oct4/Sox2* transcriptional complex (Chakravarthy *et al.*, 2008). The *Dppa2/4* duplication must have occurred late in mammalian evolution, since the opossum contains only one *Dppa2/4* gene (not shown). The neighbouring *Morc1* gene (*microrchidia*) is involved in spermatogenesis (Inoue *et al.*, 1999), and also has no orthologue in chick (Fig. 10D). Unigene-based EST expression profiling (Sayers *et al.*, 2008) shows that this gene is also expressed in the pre-implantation mouse embryo.

Other mouse EP-GRN genes not found in the chick genome are located in regions that show no syntenic conservation with chick. Therefore the only support for their being specific to mammals is that no orthologues have been found in sequence databases of chick or other non-mammalian species. Such is the case of *Dppa1* (Bortvin *et al.*, 2003) and the *miR-290* cluster (Houbaviy *et al.*, 2005; Houbaviy *et al.*, 2003) which are nonetheless considered as specific to eutherians. For other genes, additional evidence is consistent with their being specific to mammals. For example, the pluripotency marker *Rex1* (*Zfp42*), which is a retrotransposed copy of zinc finger transcription factor *YY1* and appears to be specific to placental mammals (Kim *et al.*, 2007), lies in an extremely gene-poor region of mouse chromosome 8 that is enriched in mammal-specific genes such as *Triml1*, *Adam26a*, *Adam26b* and *Adam34*. The databases contain no chick orthologue of *Nac1*, a BTB/POZ transcription factor found at the core of the EP-GRN (Kim *et al.*, 2008), but orthologues are found in fish and *Xenopus*. *Nac1* is located in a region of mouse chromosome 8 (40 Mb from *Rex1*) where all surrounding genes show the same pattern of conservation. This could indicate that the specific genomic region containing *Nac1* has been deleted in avian evolution, in line with

recent evidence showing a high level of gene loss in avians compared with other vertebrates (Hughes and Friedman, 2008).

Many genes with pivotal roles in embryonic pluripotency in the mouse are not present in the chick, suggesting that they may have arisen specifically in the mammalian lineage in connection with the establishment of the EP-GRN.

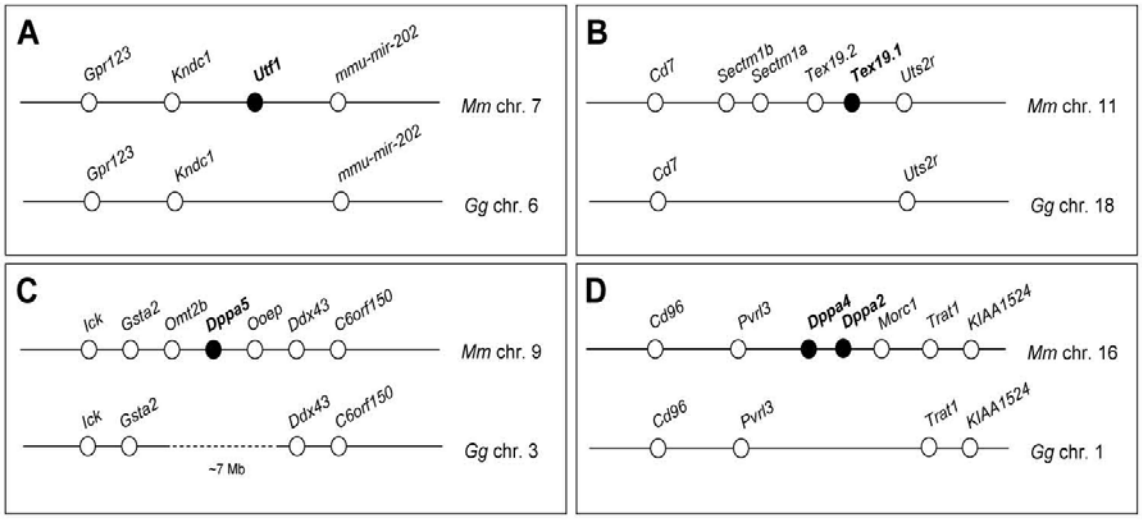


Figure 10. Lack of EP-GRN orthologues in the chick genome. Mouse *Utf1* (**A**), *Tex19.1* (**B**), *Dppa5* (**C**) and the tandem duplicates *Dppa2* and *Dppa4* (**D**) lie in regions of conserved synteny between mouse and the chick, but are absent from the latter. The mouse (*Mm*) and chick (*Gg*) chromosomes (chr.) where the genes are located are indicated. The diagrams are not shown to scale.

1.6 Core EP-GRN binding cassettes have been acquired in the mammalian lineage

Previous reports have shown that the core factors of EP-GRN bind to genomics regions regulating downstream target genes important to keep ES cells in an undifferentiated state. Moreover these analyses indicate that the core factors co-occupy a substantial portion of their targets binding in different configurations in a coordinated manner in vivo (Boyer *et al.*, 2005; Loh *et al.*, 2006). Our results above showing that in chick many components of the network have different expression domains to those in mouse or simply are not present, would predict that the regulatory targets of the core EP transcription factors were assembled de novo in mammals.

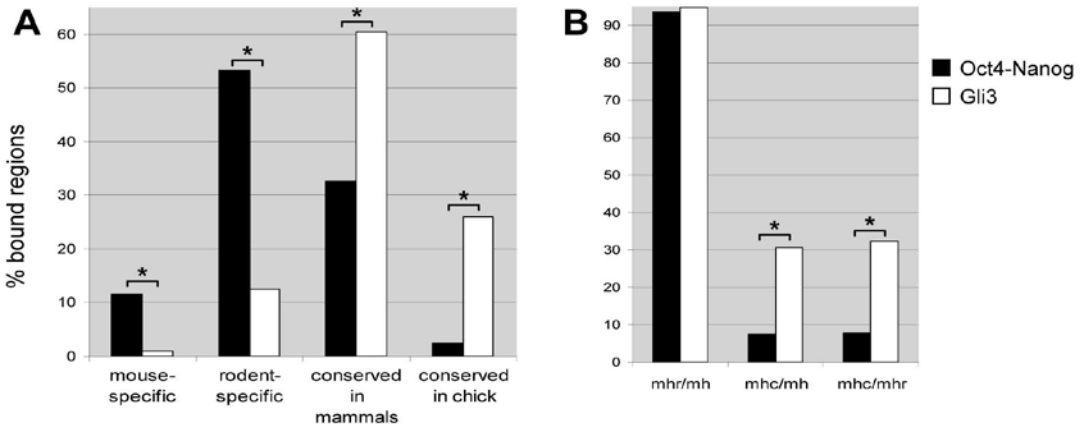
To test this, we conducted a genome-wide analysis of those regions bound by the core EP transcription factors in ES cells (Marson *et al.*, 2008). This was done by examining whole-genome alignments of several vertebrate species: human, dog, rat, mouse and chick. We reasoned that if the core of the EP-GRN network (*Oct4*, *Nanog* and *Sox2*) is indeed a mammalian novelty, genomic regions bound by these factors (most likely corresponding to cis regulatory elements) will be poorly conserved between mouse and chick. This comparison is meaningful even taking into consideration the evolutionary distance between these species and the fact that there is a considerable turnover in specific transcription factor binding sites during evolution (Odom *et al.*, 2007), as nonetheless sequence conservation can be useful in predicting approximately half of functional elements in the genome (ENCODE project consortium, 2007; Pennacchio *et al.*, 2006; Pennacchio *et al.*, 2007).

We examined the overall evolutionary conservation of 1688 non coding genomic regions bound by, at least, Oct4 and Nanog in ES cells (Marson *et al.*, 2008). Of these regions, 11.55% are not conserved between mouse and any other species, over half (53.26%) are conserved only in rodents (mouse and rat), approximately a third (32.7%) are conserved in mammals (mouse, rat, human and dog), and just 2.49% are conserved in chick (Fig. 11A). This figure is extremely low compared with prior estimates of the conservation of non-coding regulatory elements between human, rodents and chick (Hillier *et al.*, 2004). Nonetheless, to strengthen our conclusion we wished to compare the results with a similar dataset from a GRN known to control a biological process well-conserved between mouse and chick. For this, we chose the cis-regulatory circuitry underlying limb patterning, which is in many instances a prime example of the evolutionary conservation of developmental processes in vertebrates. We used the data generated in a whole-genome study where they identified the cis-regulatory network of *Gli3,* which plays a crucial role as effector of *Shh* signalling in the developing limb (Vokes *et al.*, 2008). Applying the same approach described for the regions bound by the core factors but for the 200 top-scoring genomic regions bound by GRN transcription factor Gli3, we found that 26% of these regions are conserved

between mouse and chick (Fig. 11A). It is also noteworthy that a higher proportion of the Gli3-bound regions are conserved in the other mammals analysed (60.5%, versus 32.7% for Oct4 and Nanog). This comparison shows that regions bound by Oct4 and Nanog, and therefore part of the ES pluripotency GRN, are significantly less conserved than those bound by Gli3 as part of the limb development GRN.

However, there are at least two alternative explanations for the data presented above. One possibility is that the majority of regions bound by Oct4 and Nanog in the mouse genome are cis-regulatory elements controlling mouse-specific genes that have no chick orthologues. Therefore our analysis would only detect mammalian or mouse specific EP-GRN targets. This is not the case, since we found that those genes located in the vicinity of Oct4-Nanog bound regions have a higher proportion of 1-to-1 chick orthologues than the total mouse gene set (61% versus 44%) (Fig. 11C). The second possibility is that the low conservation of Oct4-Nanog regions in the chick genome could simply be an artefact of their enrichment in mouse- or rodent-only conserved regions (Fig. 11A). To test for this, we first counted the number of Oct4-Nanog regions conserved between mouse and human, and from this set we counted those regions that were also conserved with rat or with chick. We then calculated the relative proportions of each grouping (Fig. 11C). Applying this same strategy to whole genome comparison, it has been shown that 23.29% of all non-coding regions conserved among mouse, human and rat are also conserved in chick (Prabhakar *et al.*, 2006). We found that for Oct4-Nanog regions the figure was 8.02% and for Gli3 regions it was 32.3% (Fig. 11B, C). Therefore the lack of conservation of Oct4-Nanog bound regions between mouse and chick is still significant when mouse and rodent-specific genomic regions are excluded from the analysis**.**

EP-GRN was newly assembled during the evolution of mammals, partly through the recruitment of pre-existing genes by the appearance of core EP factor response elements.



C

	All mouse genes	*Oct4-Nanog* genes
Total	28239	1591
With chick orthologues	12435	977
% with chick orthologues	44	61
Without chick orthologues	15804	614
% without chick orthologues	56	39

D

	Oct4-Nanog	*Gli3*
Conserved mouse-human (mh)	560	170
Conserved mouse-human-rat (mhr)	524	161
% conserved mh that also mhr	93.57	94.71
Conserved mouse-human-chick (mhc)	42	52
% conserved mh that also mhc	7.5	30.59
& conserved mhr that also mhc	8.02	32.30

Figure 11. Genomic regions bound by Oct4 and Nanog are poorly conserved in chick. A) A high proportion of Oct4-Nanog bound regions (black bars) are specific to mouse or conserved only in rodents, compared with regions bound by Gli3 (white bars), part of the limb development GRN. Oct4-Nanog and Gli3 bound sites are enriched in regions conserved between mammals. Very few of the Oct4-Nanog bound regions are conserved with chick, compared with those bound by Gli3 (2.49 vs. 26%). **B**) Grouping of regions bound by Oct4 and Nanog in mouse ES cells, and by Gli3 in the mouse limb that are conserved between mouse and human, and then with rat or with chick **C)** *Oct4-Nanog* set is significantly enriched in genes conserved in chick compared to the whole genome (p-value<0.0001; two-tailed Chi-square test). There are 1688 regions, giving an output of 1591 genes. We do not know if the difference comes from multiples regions for one gene, or some Ensembl IDs being wrong. **D)** The set of Oct4-Nanog bound regions has an equal proportion of conserved mammalian non-coding elements as the Gli3 set (mhr/mh), but a much lower number of regions conserved in chick (mhc/mh). mh, number of genomic regions conserved between mouse and human; mhr, regions conserved between mouse and humans that are also conserved in rat; mhc, regions conserved between mouse and humans that are also conserved in chick. Except for the mhr/mh comparison, all differences between the Oct4- Nanog and Gli3 data sets are highly significant ($p < 0.0001$ two-tailed Fisher's exact test).

1.7 Conservation and evolution of the *Fgf4* and *Sox2* EP enhancers

To further explore and refine the above scenario, we tried to identify specific cases where we could trace the appearance of a core EP-GRN regulated element. We focused on elements that have been shown to possess cis-regulatory activity in vivo in response to core EP factors. The best characterized set of such regulatory elements corresponds to Sox2/Oct4 regulated genes in which adjacent HMG and POU binding motifs mediate high transcriptional activity (Remenyi *et al.*, 2003). To date, functional HMG/POU cassettes have been characterized in detail from eight Sox2/Oct4 target genes, namely *Sox2* and *Oct4* themselves plus *Nanog*, *Lefty1*, *Fgf4*, *Fbxo15*, *Utf1* and *Dppa4* (Nakatake, 2006 #471; Chakravarthy, 2008 #469 and references therein). Only for *Fgf4* and *Sox2* we found that the genomic region containing the HMG/POU cassette shows a partial but significant degree of conservation between mouse, other mammals and chick (Fig. 12A, B). In other cases, either there is no orthologue in chick (as for *Dppa4*, *Utf1* and *Oct4* (Fig. 10 A, D and 3), or no similarity at all between mouse and chick genomic regions can be detected (as occurs for *Nanog*).

The *Fgf4* HMG/POU cassette is located in the 3' UTR of the gene, and genomic fragments carrying these sites drive reporter expression in embryonic carcinoma (EC) cells, ES cells and the ICM of E4.5 blastocysts. This expression is strictly dependant on the synergistic action of Sox2 and Oct4 (Chakravarthy *et al.*, 2008; Fraidenraich *et al.*, 1998; Yuan *et al.*, 1995), To examine the conservation of these sites and surrounding regions we aligned the 3' UTR of mouse *Fgf4* (2.3 kb) to 3 kb of genomic sequence immediately downstream from the stop codon of the coding region of *Fgf4* orthologues from other vertebrates (Fig. 12A). We found various peaks of conservation distributed along the length of the mouse 3' UTR and noticed that the Sox2/Oct4 binding site (thin blue line in Fig. 12A) was adjacent to a region conserved among mammals and chick. In this last case, sequence conservation was sufficient to anchor an alignment that contained the HMG/POU cassette (Fig. 12C). The HMG half of the cassette is identical in all species except for a single change in the 5' end in the chick sequence (A instead of C). Intriguingly, this is one of the invariable residues in all Sox2/Oct4 composite sites identified to date in mouse (Chew *et al.*, 2005). Regarding

the POU site, both chick and platypus show several nucleotide changes and a one base deletion.

The *Sox2* HMG/POU cassette localizes approximately 2.5 kb downstream of the stop codon in mouse and is contained in one of the two elements found in the vicinity of *Sox2* shown to be active in pluripotent cells (Tomioka *et al.*, 2002). We aligned 4 kb downstream from the stop codon of the mouse gene to the equivalent regions from other vertebrates. As with *Fgf4*, conservation was distributed along the length of the sequence and the Sox2/Oct4 binding site was located in a region highly conserved among all mammals, and poorly but still detectably conserved in chick and lizard (Fig. 12B). This region does not correspond to any of the multiple conserved elements previously described to act as enhancers in the early chick embryo (Uchikawa *et al.*, 2003). Closer examination showed that unlike the *Fgf4* enhancer, the *Sox2* enhancer of non-eutherian mammals contains a perfect POU site, and the HMG site contains only two changes; however, once again these affect the C residue in the first position, invariant in all other Sox2/Oct4 sites (Fig. 12D). The chick sequence is altered with respect to the mouse at multiple positions in both the HMG and the POU sites (Fig. 12D).

Based on studies of other Sox2/Oct4 elements, the nucleotide changes we observe in chick compared to the mouse *Fgf4* and *Sox2* enhancers would be predicted to abolish Oct4 and Sox2 binding and therefore enhancer activity, (Chakravarthy *et al.*, 2008; Chew *et al.*, 2005; Kuroda *et al.*, 2005; Rodda *et al.*, 2005)

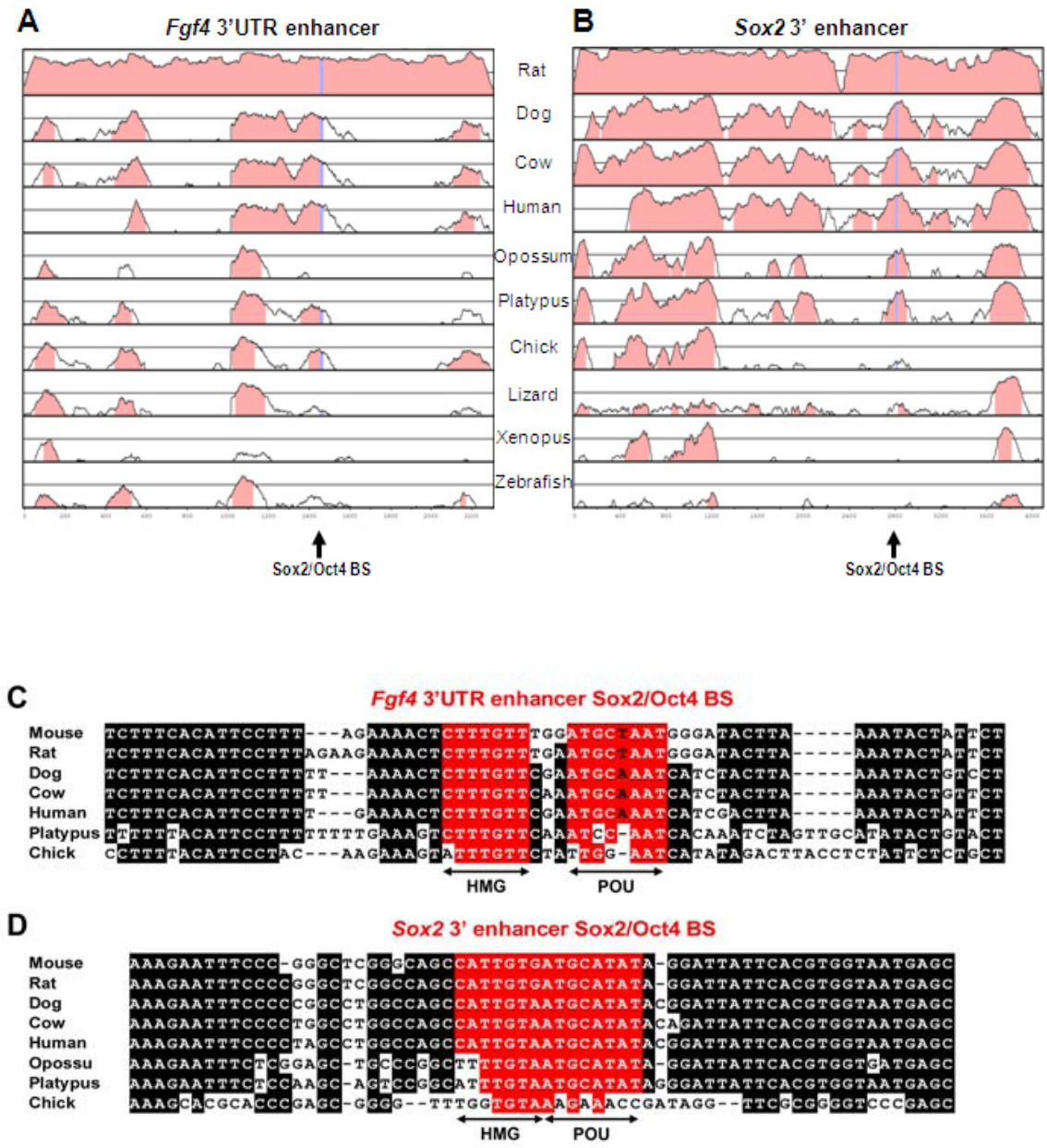


Figure 12. Conservation of the *Fgf4* and *Sox2* EP enhancers. (**A**, **B**) Vista plots comparing the 3'UTR from mouse *Fgf4* (**A**) and the 4 kb downstream of the mouse *Sox2* stop codon (**B**) with orthologous regions from other vertebrates. Regions ≥100 bp that show 60% (50% in the case of non-mammalian species) or greater sequence identity to the mouse sequence are coloured pink, and Sox2/Oct4 binding site (BS) is shown in blue. (**C**, **D**) Sequence comparison of the regions surrounding the Sox2/Oct4 BS from *Fgf4* (**C**) and *Sox2* (**D**). The HMG and POU sites are indicated below the sequence and are highlighted red. Residues identical across at least five species are shown with a black background. Dashes indicate gaps introduced to maximize the alignment. Opossum and zebrafish *Fgf4* sequences do not align in this region.

1.8 Activity of the *Fgf4* and *Sox2* EP enhancers in pluripotent cells

In order to test the enhancer activity of mouse and chick genomic regions from *Fgf4* and *Sox2,* we cloned them in an enhancer-detection vector and tested their activity in mouse ES cells and in transgenic pre-implantation embryos (Fig. 13). As a positive control, we used the previously described *Oct4* distal enhancer (*Oct4*-DE; Fig. 13B) that contains the Sox2/Oct4 binding sites and is active in both ES cells and in the inner cell mass of the blastocyst (Yeom *et al.*, 1996). As expected, both mouse fragments were able to drive expression of the reporter in ES cells, but only the *Fgf4* enhancer was active in the blastocyst transgene assay we used (Fig. 13C). On the contrary, the chick fragments showed much lower activity in ES cells, and the chick *Fgf4* fragment was not active in the blastocyst (Fig. 13D). We then proceeded to change the sequence of the Sox2/Oct4 sites from the mouse *Fgf4* and *Sox2* enhancers to that of the equivalent position in the chick sequences. This dramatically reduced expression of the reporter in ES cells (six-fold for *Fgf4*, and ten-fold for *Sox2*) (Fig. 13A) and abolished activity of the *Fgf4* enhancer in the blastocyst, confirming that bona-fide Sox2/Oct4 binding sites are necessary for enhancer activity. To test if the mouse sequence of these sites was sufficient for activity in ES cells, we mutated the corresponding base pairs in the chick fragments to those found in mouse. However, this was not sufficient to make the chick fragments active in ES cells, as the mutated fragments showed no significant changes as compared to the wild types (Fig. 13A) and the mutated chick *Fgf4* fragment was not active in the blastocyst. Experimental evidence for the case of *Sox2* has shown that changes in downstream positions that are not conserved either between chick and mouse also contribute to enhancer activity (Fig. 14) (Chakravarthy *et al.*, 2008; Tomioka *et al.*, 2002). This, together with our results, shows that sequences other than the Sox2/Oct4 cassette contained in the mouse enhancers, and not conserved with chick, are also needed for correct expression in pluripotent cells.

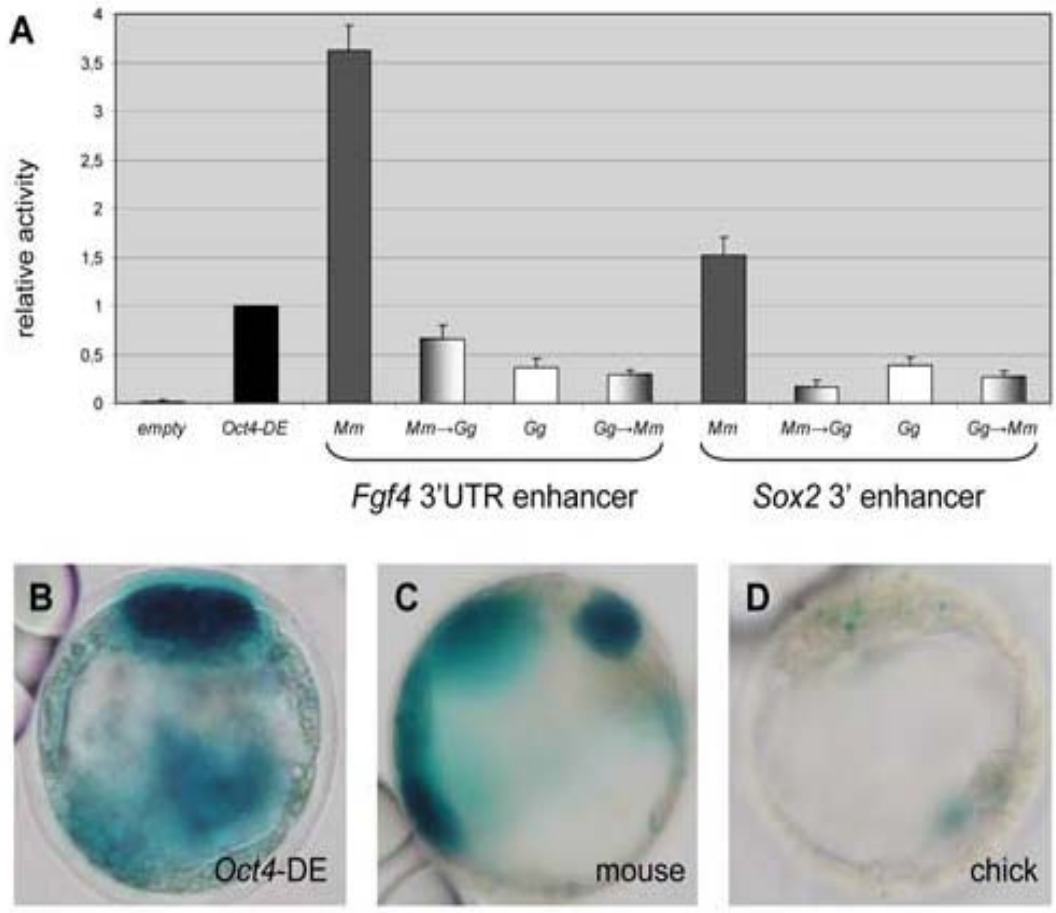


Figure 13. Activity of the *Fgf4* and *Sox2* EP enhancers. (**A**) Relative enhancer activity in ES cells of the genomic fragments from the *Fgf4* and *Sox2* locus of mouse (*Mm*) and chick (*Gg*), as well as versions where the mouse Sox2/Oct4 site were changed to the equivalent chick sequence (*Mm→Gg*), or the chick sequence was changed to include the mouse Sox2/Oct4 site (*Gg→Mm*). Transfection efficiency was normalized and expressed as relative value with respect to the activity of the *Oct4* distal enhancer (*Oct4*-DE). (**B**-**D**) Enhancer activity in transgenic mouse blastocysts of the mouse *Fgf4* 3'UTR enhancer (C) and the equivalent chick genomic region (D). In this last case, the embryo shown is representative for the unspecific punctuated pattern equal to that obtained even when the empty vector is used, as well as for the mutated versions of both the mouse and chick fragments. The activity of the mouse *Oct4*-DE in the inner cell mass is shown as a positive control (B).

A possibility that is not ruled out by the above experiments is that the nucleotide differences we observe in the chick genome correlate with changes in the binding specificities of the factors. While mouse and chick Sox2 peptides are almost identical

(Fig. 15A), ruling out drastic changes in binding specificities, mouse Oct4 and its paralogue chick Pou2-r show much greater divergence (Fig. 15B) (Lavial *et al.*, 2007).

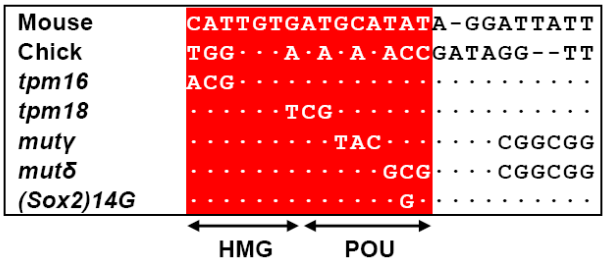


Figure 14. Alignment of the Sox2/Oct4 binding site and 3' adjacent sequences from mouse *Sox2* with chick and mutant versions that significantly affect enhancer activity. All mutations were tested by (Tomioka *et al.*, 2002), except for the (Sox2)14G mutation, which was tested by (Chakravarthy *et al.*, 2008). Dashes indicate gaps introduced to maximize the alignment; dots indicate residues identical to the mouse sequence.

A

```
Mm  MYNMMETELKPPGPQQASGGGGGGGNATAAATGGNQKNSPDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAE
Gg  MYNMMETELKPPAPQQTSGGG--TGNSNSAAN--NQKNSPDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAE

Mm  WKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMD
Gg  WKLLSEAEKRPFIDEAKRLRALHMKEHPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGTNTMTTGVGVGATLGAGVNQRMD

Mm  SYAHMNGWSNGSYSMMQEQLGYPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTPGMALG
Gg  SYAHMNGWTNGGYGMMQEQLGYPQHPGLNAHNAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTPGMALG

Mm  SMGSVVKSEASSSPPVVTSSSHSRAPCQAGDLRDMISMYLPGAEVPEPAAPSRLHMAQHYQSGPVPGTAINGTLPLSHM
Gg  SMGSVVKTESSSSPPVVTSSSHSRAPCQAGDLRDMISMYLPGAEVPEPAAPSRLHMSQHYQSAPVPGTAINGTLPLSHM
```

B

```
Mm  MAGHLASDFAFSPPPGGGDGSAGLEPGWVDPRTWLSFQGPPGGPGIGPGSEVLGISPCPPAYEFCGGMAYCGPQVGLGLV
Gg  -------------------------MHVKAKNLLRMCKWLKGLRNARGS-TWGRS---------G----------GRK

Mm  PQVGVETLQPEGQAGARVESNSEGTSSEPCADRPNAVKLEKVEPTPEESQDMKALQKELEQFAKLLKQKRITLGYTQADV
Gg  PMRSSGRLPRSADPGWGNHANRAAVVTRGISSHSPRVCLCLCQDAPTS--------EELEQFAKDLKHKRIMLGFTQADV

Mm  GLTLGVLFGKVFSQTTICRFEALQLSLKNMCKLRPLLEKWVEEADNNENLQEICKSET-LVQARKRKR-TSIENRVRWSL
Gg  GLALGTLYGKMFSQTTICRFEALQLSFKNMCKLKPLLQRWLNEAENTDNMQEMCNAEQVLAQARKRKRRTSIETNVKGTL

Mm  ETMFLKCPKPSLQQITHIANQLGLEKDVVRVWFCNRRQKGKRSSIEYSQREE---YEATGTPFPGGAVSFPLPPGPHFGT
Gg  ESFFRKCVKPSPQEISQIAEDLNLDKDVVRVWFCNRRQKGKRLLLPFGNESEGVMYDMNQSLVPPG---LPIP----VTS

Mm  PGYGSPHFTTLYSVPFPEGEAFPSVPVTALGSPMHSN
Gg  QGYSLAPSPPVYMPPFHKAEMFPPPLQPGISMNNSSH
```

Figure 15. Alignment of the predicted proteins of *Sox2* and *Oct4* genes. (**A**) Mouse and chick Sox2 proteins show overall similarity of 92% (**B**) Mouse Oct4 and chick Pou2-r proteins only 37%. Mm, *Mus musculus*; Gg, *Gallus gallus*.

We therefore tested the transcriptional activity of mouse *Oct4* and chick *Pou2-r* on the different mouse pluripotency enhancers and equivalent genomic regions from chick. In

a first set of experiments we use the mouse ES cell line ZHBTc4, where the addition of tetracycline to the culture leads to silencing of *Oct4* (Niwa *et al.*, 2000). Upon transfection of mouse *Oct4* or chick *Pou2-r* and in the absence of tetracycline, the activity of *Oct4*-DE is increased due to the effects of the exogenous Oct4 and Pou2-r proteins. Upon silencing with tetracycline, endogenous *Oct4* is repressed and consequently *Oct4*-DE activity is lost. However when mouse Oct4 or chick Pou2-r are overexpressed both are equally able to bind to *Oct4*-DE and activate the enhancer, rescuing its activity (Fig. 16A).

The E14 ES cell line, more robust in reporter expression assays, was used to test the response of *Fgf4* and *Sox2* chick fragments to Pou2-r overexpression. Even though these cells do not have an inducible-repression system for *Oct4*, the experiment above demonstrated we are able to detect differences in enhancer activity in the presence of endogenous *Oct4* (Fig 16A tet-). The overexpression of chick Pou2-r did not have any effect on chick genomic fragments from *Fgf4* or *Sox2* or on mutated version of the mouse enhancers where nucleotide residues in the binding sites had been changed to those present in chick (Fig. 16B). Therefore Pou2-r is not able to bind them, confirming the chick fragments do not contain cryptic POU binding sites that respond specifically to chick Pou2-r.

The binding specificities of chick Sox2 and Pou2-r are conserved with mouse. However, regions in the chick genome corresponding to Sox2/Oct4 responsive elements in the mouse genome cannot be bound by these factors because they lack critical sequences. The number of nucleotide differences between the chick and mouse sequences is small enough to have arisen by point mutation and insertion. Thus, modest changes led to the appearance of novel regulatory elements under the control of *Sox2* and *Oct4*, as part of the emerging EP-GRN during mammalian evolution.

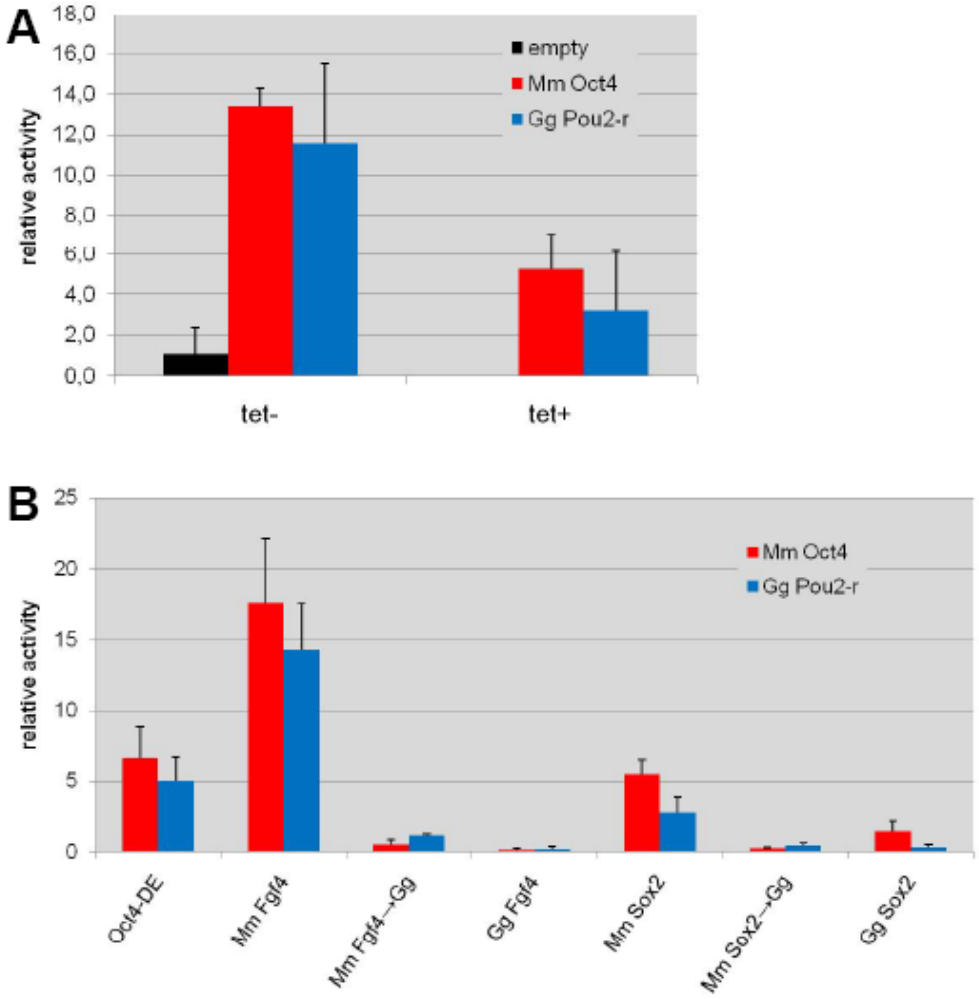


Figure 16. Overexpression of mouse Oct4 and chick Pou2-r have similar effects on pluripotency enhancers in ES cells. (**A**) The activity of the *Oct4* distal enhancer (*Oct4*-DE) in ZHBTc4 ES cells is increased by the overexpression of both mouse Oct4 and chick Pou2-r in similar degree (tet-). Shut-down of endogenous *Oct4* expression by the addition of tetracycline (tet+) abolishes the activity of the *Oct4*-DE (empty), but enhancer activity is recovered when mouse Oct4 or chick Pou2-r are overexpressed. Relative activity of the enhancer element was determined by the reporter mRFP expression (**B**) Mouse *Fgf4* and *Sox2* pluripotency enhancers (*Mm Fgf4, Mm Sox2*) respond in a similar degree to overexpression of mouse *Oct4* or chick *Pou2-r* in E14 ES cells. Neither the chick genomic regions equivalent to the mouse enhancers (*Gg Fgf4, Gg Sox2*) nor the mouse enhancers where the Sox2/Oct4 site was changed to the equivalent chick sequence (*Mm Fgf4→Gg, Mm Sox2→Gg*) showed any increased response to overexpression of chick Pou2-r. The activity of the *Oct4*-DE is shown as a control. Relative activity of enhancer elements was determined by the *lacZ* reporter expression. Standard deviation bars are shown.

2. FUNCTIONAL ANALYSIS OF VERTEBRATE *NANOG* FACTORS

2.1 Conservation and divergence of Nanog proteins

Nanog is a divergent homedomain protein found in pluripotent mammalian embryo cells, in derived ES cells and in developing germ cells. *Nanog* mediates the naïve epiblast pluripotent ground state and is indispensable for the formation of germ cells (Chambers *et al.*, 2007; Silva *et al.*, 2009). *Nanog* orthologues have been identified in several vertebrates (Canon *et al.*, 2006; Dixon *et al.*, 2010; Lavial *et al.*, 2007) and the alignment of their predicted proteins has revealed intriguing differences with mouse Nanog, particularly in a region crucial for its biological function (Mullin *et al.*, 2008; Wang *et al.*, 2008). In eutherian mammals this region contains repeated motifs in which every fifth amino acid is tryptophan, and this region is essential for formation of a functional dimeric Nanog protein. These motifs are present in ten copies in mouse and eight copies in human. However, opossum Nanog contains only two copies, and the chick orthologue contains none (Fig. 17). To test the functional significance of this difference, we compared vertebrate Nanog proteins in functional assays in mouse ES cells.

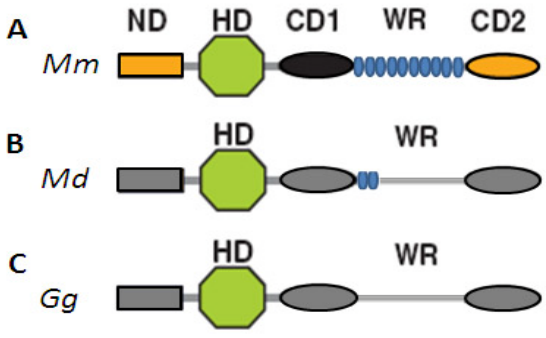


Figure 17. Schematic representation of the predicted Nanog protein structures in (A) mouse, (B) opossum and (C) chick. The conservation is the highest in the homedomain. The five amino acid tryptophan repeat motif is present in ten copies in mouse Nanog (A, blue), two copies in opossum (B, blue) and none in chick (C, blue). Domains that are less than 50% similar to mouse Nanog are coloured grey. ND, N-terminal domain; HD, homedomain; CD1, CD2, C-terminal domain 1 and 2; WR, dimerization or tryptophan repeat domain; Mm, *Mus musculus*; Md, *Monodelphis domestica*; Gg, *Gallus gallus. Modified from* (Chambers and Tomlinson, 2009).

2.1.1 Functional equivalence of vertebrate Nanog proteins

Mouse *Nanog* function was originally described in mouse ES cells; Nanog over-expression in these cells can bypass the requirement of LIF and maintain ES pluripotency (Chambers *et al.*, 2003). Therefore ES cells are a good system to test the function of chick *Nanog*. We also analysed the *Nanog* orthologue of the marsupial opossum. *Nanog* null mouse ES cells (Chambers *et al.*, 2007) were stably transfected with vectors containing mouse, opossum or chick *Nanog.* Since these cells lack endogenous mouse *Nanog*, we could explore the extent to which opossum or chick Nanog overexpression sustains self-renewal after LIF withdrawal. Pluripotency of the ES colonies after transfection was assessed by alkaline phosphatase staining, and colonies were classified as undifferentiated, mixed or fully differentiated (Fig. 18). Both orthologues were able to maintain colonies with a normal ES morphology and the stable clones retained self-renewal and pluripotency after several passages.

Despite the sequence divergence between mouse Nanog and the Nanog orthologues from opossum and chick, these proteins are functionally equivalent in mouse ES cells.

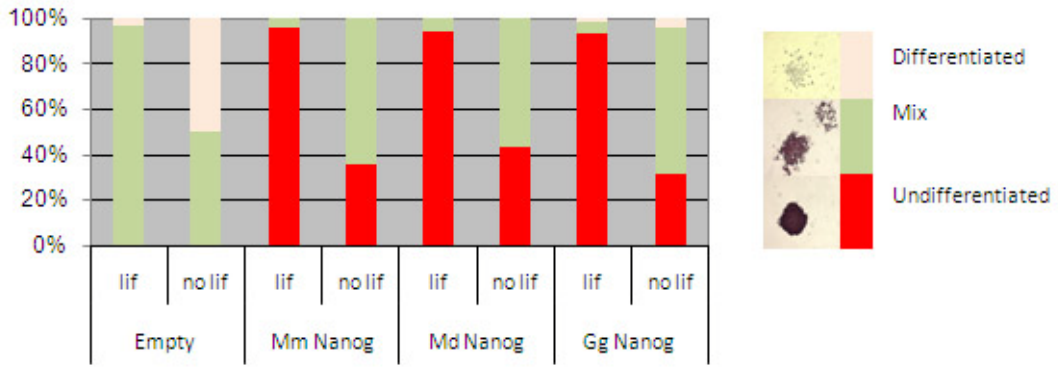


Figure 18.Functional equivalence of mouse, opossum and chick Nanog proteins in *Nanog*-null ES cells. Percentages of three categories of colonies stained for alkaline phosphatase: differentiated, mixed or undifferentiated. In the absence of *Nanog* (empty) only mixed or differentiated colonies are found in the no LIF condition. In the absence of LIF, expression of *Nanog* sustained the undifferentiated state to a similar extent irrespective of the species of the protein (34% of colonies for mouse *Nanog*, 42% for opossum, and 30% for chick).

2.2 Downstream targets of *Nanog* in vertebrates

Functional equivalence in tissue culture assays indicate that chick *Nanog* is able to bind to mouse *Nanog* targets in ES cells but it does not imply it has the same functioning or networking in vivo. In fact, the different *Nanog* expression patterns (Canon *et al.*, 2006) and our previous results on EP-GRN would suggest that the range of *Nanog* targets differs between mouse and chick. We aimed to elucidate the differences by comparing the actions of overexpressed Nanog in both species in vivo.

2.2.1 Nanog overexpression in the chick embryo

Chick Nanog was overexpressed in chick embryos by electroporation. This method of transient transgenesis allows efficient overexpression of a cDNA cloned into a vector containing a strong promoter. By applying an electric field, the construct can be delivered into the cells of the embryo. In this manner, we overexpressed chick Nanog in the epiblast cells of HH4 chick embryos, the earliest stage at which this approach is technically possible. In all cases, empty vector (control) or *Nanog* was co-electroporated with a GFP-expressing construct to mark the area of electroporation (Fig. 19, 20, 21).

We next examined the expression of *Pou2-r* and *Sox2*, the homologues of the core EP-GRN factors that are upregulated by *Nanog* (Boyer *et al.*, 2005; Wang *et al.*, 2006) We also analysed the orthologues of *Eomes* and *Cdx2*, which specify extraembryonic fate in mouse development and are repressed by *Nanog* in the embryonic lineage and in ES cells (Chen *et al.*, 2009).

In control embryos *Pou2-r* expression was detected as described (Lavial *et al.*, 2007), while in embryos overexpressing Nanog, *Pou2-r* was upregulated (Fig. 19A). In mouse ES cells even though *Oct4* expression is activated by Nanog it also repressed by Oct4 itself. Therefore Nanog overexpression fails to increase Oct4 beyond the steady-state concentration (Pan *et al.*, 2006). These results suggest that in chick this *Oct4* autoregulatory loop might not be present.

Conversely, Nanog overexpression did not alter the *Sox2* expression pattern in the neural tube, cephalic placodes or its derivates (Rex *et al.*, 1997; Uchikawa *et al.*, 2003)

(Fig. 19B), despite the fact that Nanog positively regulates *Sox2* expression in mouse ES cells (Boyer *et al.*, 2005). Since *Sox3* is expressed in part of the epiblast at stage EGK-X but *Sox2* is not, we therefore examined *Sox3* expression, in order to exclude an exchange in the response of these two factors to Nanog. Nanog overexpression had no effect on the *Sox3* expression pattern (Fig. 19C) (Rex *et al.*, 1997).

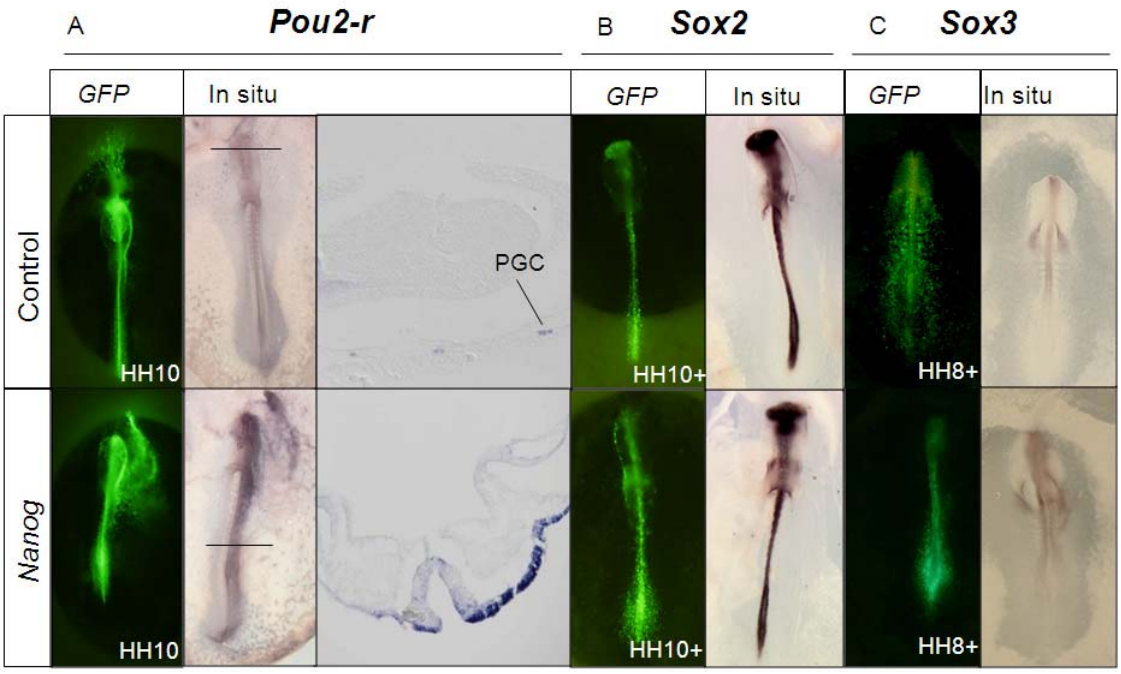


Figure 19. Nanog overexpression causes upregulation of *Pou2-r* but does not change *Sox2* or *Sox3* expression. (A) In controls, *Pou2*-r is detected in PGCs and in neural tissue; in Nanog overexpressing embryos (*Nanog*) *Pou2*-r is upregulated in the electroporation area. **B)** In controls, *Sox2* is detected in the neural tube and the cephalic placodes and its derivates; in Nanog overexpressing embryos the expression pattern does not change. **C)** *Sox3* expression in controls overlaps with that of Sox2, and is similarly unaffected in Nanog overexpressing embryos.

Nanog overexpression also had no effect on *Cdx2* expression. It might have been predicted than Nanog overexpression would downregulate *Cdx2*, in analogy with results from mouse ES cells (Chen *et al.*, 2009). However, *Cdx2* expression was unaltered, even when we carefully electroporated the embryo's caudal part, where *Cdx2* is normally expressed (Marom *et al.*, 1997) (Fig. 20A).

Interestingly, when we examined *Eomes*, not only was this factor not repressed by Nanog; its expression was upregulated (Fig. 20B). Although surprising, this result is in

line with a recent report from our group showing that *Eomes* is expressed in chick embryos in a novel domain not conserved with mouse, the PGCs (Pernaute *et al.*, 2010). Since *Nanog* is also expressed in PGCs (Canon *et al.*, 2006) the interrelationship between these two genes must have changed since the separation of birds and mammals.

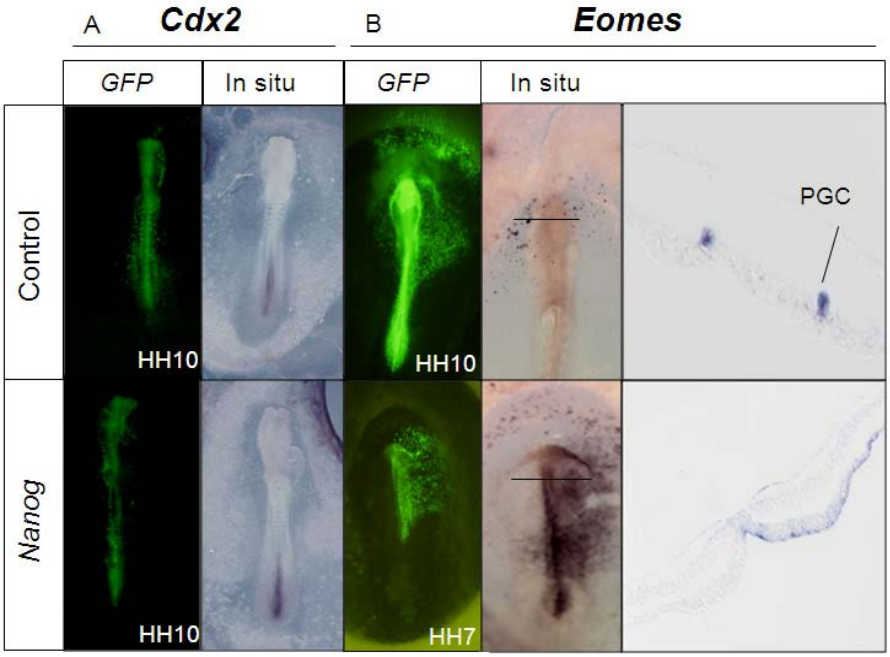


Figure 20. Nanog overexpression does not change *Cdx2* expression but upregulates *Eomes.* **(A)** *Cdx2* is expressed in the caudal part of the primitive streak and its expression is unaltered by Nanog overexpression **(B)** In control embryos, *Eomes* expression is detected in the PGCs. *Eomes* expression is strongly upregulated by Nanog overexpression in the area of electroporation.

A possible explanation to the previous results is that *Nanog* drives a general reprogramming, changing epiblast fate towards PGCs. If so, overexpression of Nanog would lead to a general upregulation of PGC markers such as *Pou2-r* and *Eomes*. We therefore examined the effect of Nanog overexpression on *Vasa*, one of the earliest and most specific germ cell markers across vertebrates (Tsunekawa *et al.*, 2000). We did not find any change in Vasa expression, ruling out an overall reprogramming of Nanog overexpressing epiblast cells to a PGC fate (Fig. 21).

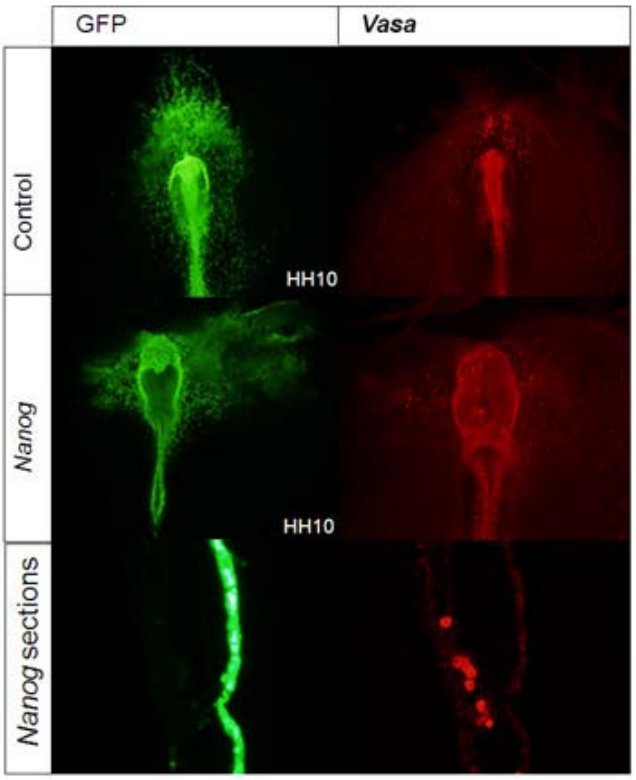


Figure 21. Nanog overexpression does not change Vasa expression. Immunohistochemistry with Vasa antibody stains only PGCs in both control and Nanog overexpressing embryos, as revealed by examining sections of Nanog overexpressing embryos, in which the PGCs are brighter and clearly visible above background.

2.2.2 Nanog overexpression in the mouse embryo

We overexpressed Nanog in the mouse embryo by using a doxycycline-dependent expression system (Hochedlinger *et al.*, 2005). Overexpression of Nanog was induced by giving doxycycline to pregnant females during three days immediately before embryo extraction at E9.5, a stage equivalent to that studied in chick (HH10). Before analysing other core factors and TE markers, we confirmed Nanog overexpression by comparing embryos positive for both transgenic alleles (Tg.TetOP-*Nanog* and ROSA26::rtTA) with embryos that did not have both alleles. We refer to these embryos as Nanog and controls respectively. A robust expression of *Nanog* was obtained throughout Nanog embryos (Fig. 22A, B).

Nanog overexpression did not upregulate *Oct4* or *Sox2* to levels detectable by in situ hybridization (Fig 22). Their expression patterns (Avilion *et al.*, 2003; Scholer *et al.*, 1990) were maintained in Nanog overexpressing embryos (Nanog).

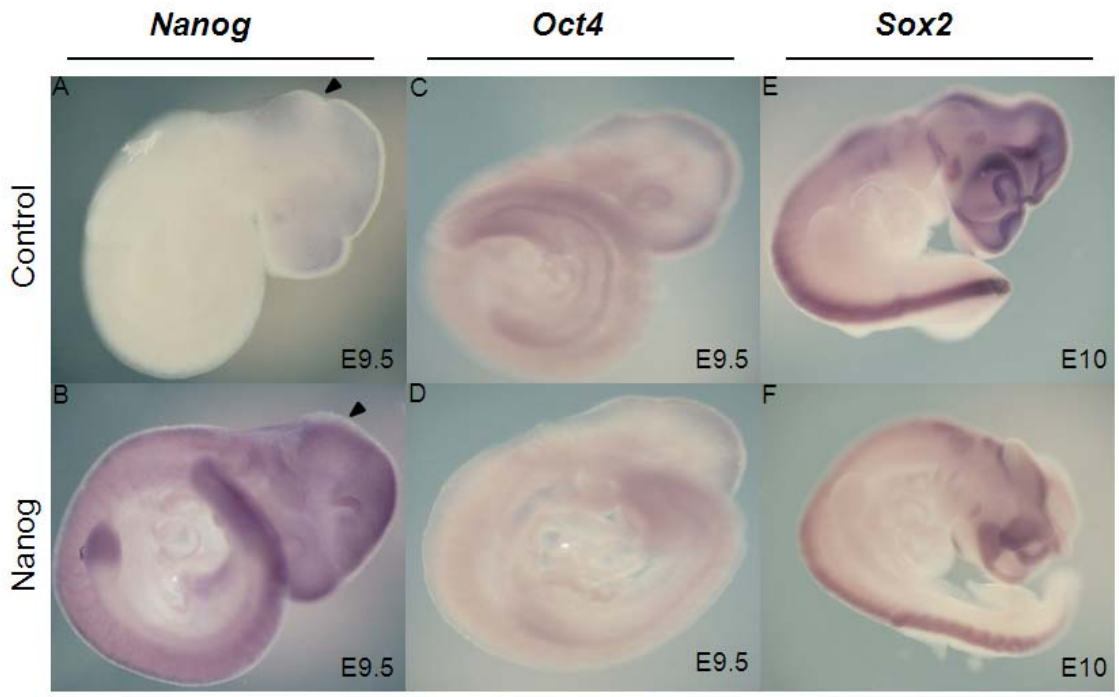


Figure 22. Nanog overexpresion does not change *Oct4* or *Sox2* expression in the mouse embryo. (A, B) In Nanog embryos *Nanog* is ectopically expressed (B) Arrowheads mark the location of the mid-hindbrain constriction **(C, D)** *Oct4* expression is not upregulated in Nanog embryos (D). **(E, F)** In control embryos, *Sox2* is detected throughout the brain, neural tube, sensory placodes, branchial arches and gut endoderm (E), a pattern unaltered in Nanog embryos (F).

Expression of the TE marker *Cdx2* in the posterior axial expression domain (Beck *et al.*, 1995) was not downregulated in Nanog overexpressing embryos (Fig. 23 A, B).

Eomes expression is detected in the anterior neural domain in control embryos as described (Ciruna and Rossant, 1999) (Fig. 23 C, E), while it appears to be disrupted in the Nanog overexpressing embryos (Fig. 23 D, F).

Closer examination of these embryos revealed they had a reproducible cranio-facial phenotype with a variety of morphological defects, among the most apparent being a lack of constriction between the midbrain and hindbrain (Figure 22 A, B).

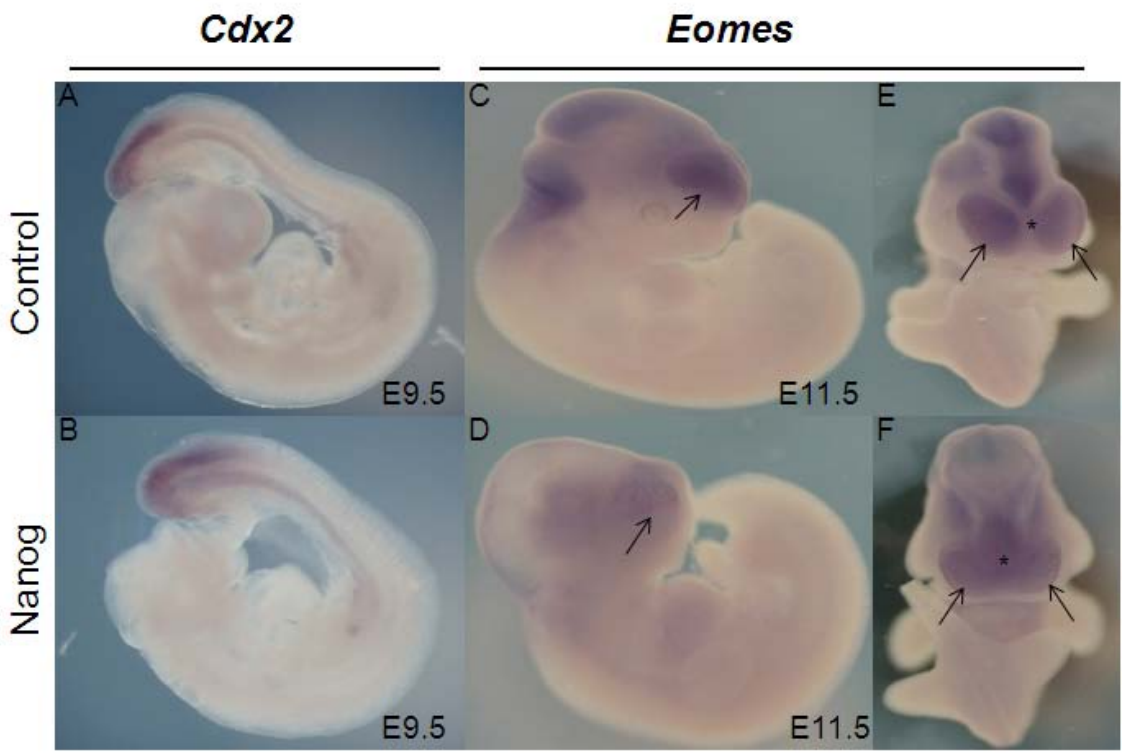


Figure 23. Nanog overexpression does not change *Cdx2* expression but alters *Eomes* (A, B) In control embryos (A) *Cdx2* expression is detected in the caudal pole, the posterior neural tube and the posterior gut endoderm, an expression pattern also found in Nanog embryos (B). **(C-F)** In the control embryos (C, D) *Eomes* expression is strong in the forebrain, where it is detected in the superficial layer of the telencephalic vesicles (arrows), and is excluded from the dorsal midline (asterisk). In Nanog embryos, expression is also found in the telencephalic vesicles but the pattern appears to be disrupted by the morphological defects.

This observation led us to examine *Fgf8,* a key organizer in the development of the anterior nervous system, which is expressed in this constriction (called isthmus) (Sato and Joyner, 2009). Although *Fgf8* is also expressed in the pharyngeal arch, developing limbs and tail bud (Crossley *et al.*, 1996), in Nanog overexpressing embryos we found that the expression was altered exclusively in the area where the constriction normally develops. The expression was diffuse and extended anteriorly (Fig. 24 B, D), while in control embryos *Fgf8* expression was restricted to a sharp band marking the isthmus (Fig. 24 A, C).

We wondered if *Fgf8* expression might also be affected in chick. To test this we examined Nanog overexpressing chick embryos for expression of *Fgf8*. The chick orthologue is expressed in territories equivalent to those in mouse (Crossley *et al.*, 1996). *Fgf8* expression was altered in the anterior region and might also be disrupted in the tail bud (Fig. 25).

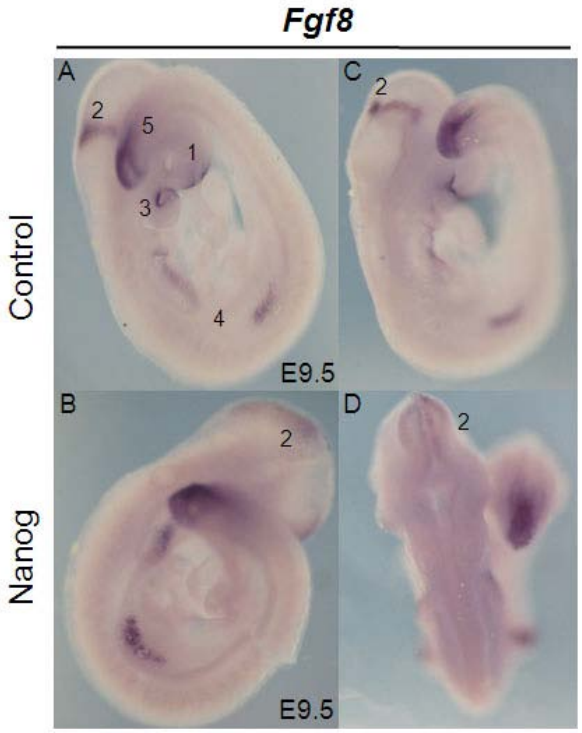


Figure 24. *Fgf8* expression pattern is disrupted by Nanog overexpression in mouse embryos. (A, C) In control embryos *Fgf8* expression is found in the prospective forebrain (1), the midbrain-hindbrain junction—in a constriction called the isthmus (2), the pharyngeal arches (3), developing limbs (4) and tail bud (5). In Nanog embryos (B, D) the sharp band of Fgf8 expression in the isthmus disappears, but the other expression domains are not affected.

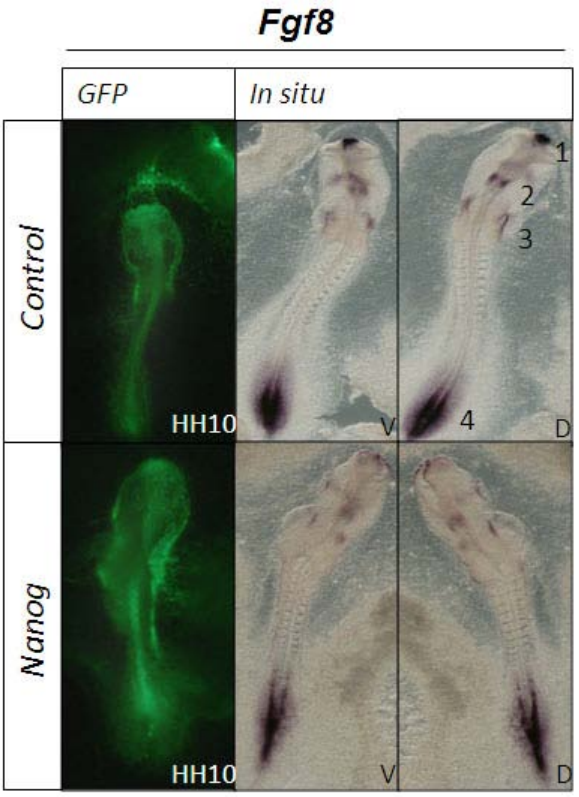


Figure 25. *Fgf8* expression pattern is disrupted by Nanog overexpression in chick embryos. In control embryos *Fgf8* is detected in cells surrounding the anterior neuropore in the forebrain (1); neuroepithelial cells in the isthmus (2); the branquial arch region (3); and the tail bud (4). In Nanog overexpressing embryos, *Fgf8* expression is disrupted in the anterior region, where embryos exhibit morphological defects, and also appears to be disrupted in the tail bud. D, dorsal view; V, ventral view.

2.2.3 Global gene expression analysis

To extend our analysis we examined gene expression in Nanog overexpressing embryos (E9.5) by microarray. 476 genes were differentially expressed in the Nanog overexpressing embryos compared with controls (corrected P-value cut-off set at 0.05). Of these genes, 292 were downregulated and 184 were upregulated.

Among the altered genes we distinguished two clear groups: one set of genes involved in nervous system development and function (87 genes) (Fig. 26), in line with our previous results showing morphological defects in the developing nervous system; and a second set of genes involved in haematological development (79 genes) (Fig. 27). Within both sets, most genes were downregulated.

The set of downregulated genes associated with nervous system development included *Otx2 and Sox2* (corrected P-value 0.02638322 and 0.01125896 respectively). Interestingly, one of the upregulated genes in this set was *Fgf8* (corrected P-value 0.047891). *Otx2* function is related to *Fgf8*, since it is involved in the correct positioning of the mid-hindbrain constriction, limiting *Fgf8* expression anteriorly (Li and Joyner, 2001). *Fgf8* might be directly upregulated by Nanog and in turn repress Otx2, a relationship already demonstrated by (Liu *et al.*, 1999). Another possibility is that *Nanog* directly represses *Otx2*, expanding *Fgf8* expression into the dorsal midbrain, as occurs in *Otx2*+/- mice (Martinez-Barbera *et al.*, 2001).

Published evidence would predict that *Sox2* expression would be upregulated (Boyer *et al.*, 2005). Instead, we observed no change by *in situ* but a repression in the microarray. This repression might be a consequence of the neural tube defects in Nanog overexpressing embryos, since *Sox2* has functions in nervous development (Miyagi *et al.*, 2008)

Arx	*Fezf1*	*Hes5*	*Lhx2*	*Nkx2-1*	*Pou3f1*
Ascl1	*Fezf2*	*Heyl*	*Lhx5*	*Nkx2-2*	*Pou3f3*
Barhl1	*Fgf15*	*Id4*	*Lmo3*	*Nkx2-4*	*Scube2*
Dbx2	*Foxg1*	*Irx1*	*Lmx1a*	*Nkx6-2*	*Six6*
Dlx4	*Fzd2*	*Irx2*	*Neurod4*	*Otx2*	*Sox2*
Dmbx1	*Hes2*	*Lfng*	*Neurog2*	*Pax6*	*Wnt7b*
					Wnt8b

Figure 26. List of representative genes involved in nervous system development and function downregulated by Nanog. The genes were associated with their function in developing nervous system by using Mouse genome informatics (www.informatics.jax.org) and Ingenuity® Systems analysis (www.ingenuity.com).

In the set of genes related to haematological system development, we identified certain genes, like *Gata1* and *Klf1*, which have central roles in erythroid differentiation (Fig. 27) (Tallack *et al.*).

A first examination did not associate upregulated genes with specific functional categories. However, in addition to *Fgf8*, two pluripotency genes were presented: *FoxD3*, which acts downstream of *Oct4*, *Sox2* and *Nanog*; and *Oct4* itself, although with a corrected P-value just below the cut-off (0.055125). Nanog overexpression might not upregulate significantly *Oct4* because this gene has a negative feedback loop to regulate its own quantity at a rigid concentration (Pan *et al.*, 2006) necessary to prevent ES cells differentiation (Niwa *et al.*, 2000).

Since we cannot distinguish between *Nanog* direct and indirect target, we mapped the vicinities of differentially expressed genes for the presence of Nanog binding sites characterized in ChIPseq analysis in ES cells (Chen *et al.*, 2008; Marson *et al.*, 2008). We found Nanog binding sites in *Sox2* and *Oct4,* as expected (Boyer *et al.*, 2005; Wang *et al.*, 2006), but also found sites in other differentially expressed genes, including *Fgf8* (a binding site not conserved in chick), *Otx2, Pax6, Dbx2, Lhx2* and *Ascl1.*

Target responsiveness to Nanog overexpression differs between chick and mouse embryos. The subnetwork downstream of *Nanog* might have changed during evolution, with new regulatory interactions appearing in the mammalian lineage, such as *Oct4* autoregulation or *Eomes* repression by *Nanog*. Neural and haematopoietic alterations observed in Nanog overexpressing embryos suggest this gene, in addition to its pluripotency function, acts as a repressor of differentiation in embryonic development.

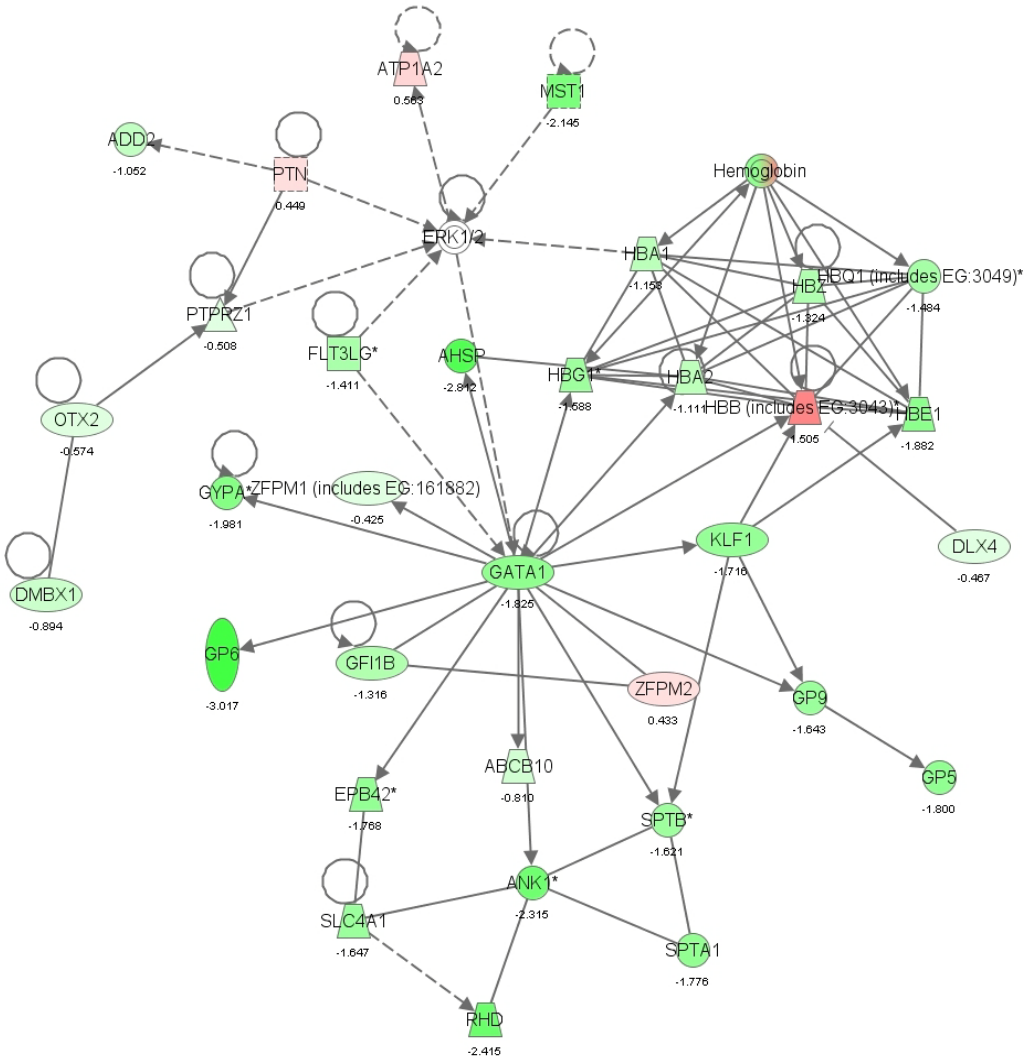


Figure 27. Hematological system development network. Genes are represented as nodes, and the biological relationship between two nodes is represented as an edge (line). Lines represent direct interactions and dotted lines indicate indirect interactions. The intensity of the node color indicates the degree of up- (red) or down-(green) regulation and is proportional to the fold difference in gene expression. Numbers underlying each molecule indicate the fold change logarithm (logFC). Nodes are displayed using various shapes that represent the functional class of the gene product. (Ingenuity® Systems, www.ingenuity.com).

DISCUSSION

Pluripotency is the ability of a cell to generate all the cell types of an adult organism, a feature shared by early embryonic cells in the developing embryo and ES cells. Pluripotency is controlled in mammals by the action of a network of genes whose core factors are *Oct4*, *Sox2* and *Nanog*. Despite significant advances in the field, little is known about the evolutionary history of this gene network and if it plays similar roles in non-mammalian vertebrates.

A recent study argues for conservation of pluripotency and its genetic control in non-mammalian vertebrates (Dixon *et al.*, 2010). These authors described a homologue of the core network in axolotl and how it functions in mouse ES cells. They suggest, similarly to previous studies in chick and *Xenopus* (Lavial *et al.*, 2007 ; Morrison and Brickman, 2006), that pluripotency factors are functionally equivalent among vertebrates. Our experiments confirm that opossum and chick *Nanog* are also able to functionally replace mouse *Nanog* in ES cells. Together, these experiments show that the homologues of the core factors are able to bind to mouse EP targets in ES cells, suggesting the core factor binding specificities might be conserved between mouse and other vertebrates. However, this does not necessarily imply conservation of the pluripotent state in these organisms, since the function and networking of these factors *in vivo* might differ.

Recent comparative genome-wide studies between different mammalian species in fact point to a high degree of plasticity of the EP-GRN (Kunarso *et al.*, 2010; Xie *et al.*, 2010). The EP-GRNs of mammals share core components indispensable for pluripotency, and the variability is located mainly in the peripheral components. The authors described that these components underwent substantial alterations through single nucleotide mutations, leading to turnover of transcription binding sites and insertion of cis-regulatory modules partly mediated by transposons. The outcome of these changes was a rewiring of the EP-GRNs in mammals.

Our results comparing mouse and chick provide further insights into the evolution of embryonic pluripotency and suggest that the EP-GRN is a mammalian novelty. The EP-GRN arose not only from the appearance of new pluripotency genes but also by co-opting and duplicating existing ones and by establishing new regulatory inputs between them. We found that some EP-GRN genes are specific to mammals but that

the core pluripotency factors were already present in avians and other vertebrates, even though they might have different roles to their homologues in the mouse embryo. Furthermore, the way the pluripotency factors are connected in the EP-GRN appeared in mammals, and the network subsequently underwent further rewiring (Kunarso *et al.*, 2010; Xie *et al.*, 2010).

Our results from Nanog overexpression in chick and mouse embryos also suggest that many pre-existing genes have been rewired into the mammalian EP-GRN by the appearance of new regulatory interactions. This is supported by a study comparing zebrafish and mouse *Pou5* transcriptional networks, which suggests that *Pou5* subnetworks were co-opted in mammals for additional novel functions based on new interacting partners or feedback loops (Onichtchouk *et al.*). For example, the *Oct4* autoregulatory mechanism characteristic of mammals is not present in zebrafish (Onichtchouk *et al.*) or in chick, as suggested by our results from Nanog overexpression. We also found that in the mammalian lineage *Nanog* became a repressor of *Eomes* (Fig 1).

Since our conclusions stem mainly from comparison with mouse, a potential concern might be the degree of conservation of the EP-GRN among mammals. The most detailed comparisons among mammals carried out so far have been between mouse and human ES cells. These studies show that the core factors (Boyer *et al.*, 2005) and many other EP-GRN genes (Kim *et al.*, 2005; Kuntz *et al.*, 2008; Monk *et al.*, 2008) are conserved in both mammalian species. However, striking differences have been found between human and mouse ES cells in their morphology and culture requirements (Pera and Trounson, 2004), targets of EP-GRN core factors (Boyer *et al.*, 2005; Loh *et al.*, 2006), and critical signalling networks and gene expression profiles (Ginis *et al.*, 2004). These differences might indicate that mouse and human ES cells are not equivalent embryonic cell populations, and therefore are not directly comparable. Indeed, human ES cells share more characteristics with mouse epiblast-derived stem cells (EpiS) than with mouse ES cells (Brons *et al.*, 2007; Tesar *et al.*, 2007). Therefore, human ES cells might not be equivalent to mouse ES cells but to mouse EpiS cells.

1. Early patterning in the chick embryo

We examined the pre-gastrulation chick embryo for the expression of homologues of important mouse EP-GRN genes. The core pluripotency factors are present across amniotes but their expression patterns in the early chick embryo are different from those found in mouse. Chick *Nanog* is expressed only in primordial germ cells of the pregastrulation chick embryo (Canon *et al.*, 2006); the putative chick *Oct4* homologue ubiquitously expressed in the epiblast at stage EGK-XI (Fig. 1) (Lavial *et al.*, 2007) is not a true orthologue of mouse *Oct4*, but a paralogue. Our analysis shows that avians lack a true *Oct4* orthologue. In fact the complex duplicative history of *Oct4* that we found in vertebrates has been confirmed by the presence of *Pou2*-r and fully functional *Oct4*-r genes in a monotreme, the platypus, (Niwa *et al.*, 2008) and in a marsupial, the tammar wallaby (Frankenberg *et al.*, 2010). Niwa *et al.* suggest that the *Oct4* gene found today in eutherian mammals is the result of a duplication event which occurred during early mammalian divergence. Furthermore these authors propose a causal relationship between the duplication and the appearance of a new function of *Oct4* in the early mammalian embryo: the reciprocal inhibition of *Oct4* and *Cdx2*. However, our finding that the *Oct4* genomic region may have been specifically lost in avians, together with our identification of an *Oct4* orthologue in the lizard genome, date the duplication of the *Oct4*-related genes before the divergence of avians and mammals, as recently suggested by (Frankenberg *et al.*, 2010). These findings refute any relationship between the duplication event and the acquisition of the new interaction between *Oct4* and *Cdx2*.

In addition we found that other transcription factors such as *Sox2* and *FoxD3* and the signalling molecules *Fgf4* and *Nodal* are not expressed until post-gastrulation stages, in contrast to their mouse orthologues which are expressed in the blastocyst ICM and epiblast (Fig. 1) (Avilion *et al.*, 2003; Hanna *et al.*, 2002; Niswander and Martin, 1992; Takaoka *et al.*, 2006).

It could be argued that, in some cases, paralogues might be redundant. This could be the case for *Sox1* and *Sox3*, paralogues of *Sox2*. However, *Sox1* is not expressed in the epiblast, and *Sox3* is not expressed in the whole epiblast (Wilson *et al.*, 2001) and the

significance of its epiblast expression is difficult to judge. First, *Sox3* is not able to fully replace *Sox2* in the generation of iPS cells (Nakagawa *et al.*, 2008); and second, Sox3 is not expressed in the early mouse embryo until post-implantation stages (www.informatics.jax.org). Therefore, if *Sox3* were the partner of *Pou2*-r in chick, this would imply that the mammalian lineage had lost early *Sox3* expression concomitantly with the gain of pluripotent expression of *Sox2*. Thus, even in the case of an exchange of function between *Sox2* and *Sox3*, this would have involved dramatic changes in the regulation and function of these genes and the network between avians and mammals.

We did not examine the expression of all these genes at earlier stages of embryonic development inside the chick oviduct, when they might exert a role in embryonic pluripotency and lineage specification. However, such a role is unlikely, since at these stages the zygotic genome has still not activated, and these genes could therefore only be expressed maternally: For these genes to exert their differential functions, there would have to be mechanisms for asymmetrically distributing maternal transcripts between the different territories of the embryo or for permitting their translation only in specific regions. It is true that maternal expression of the zebrafish *Pou5f1*-related gene *pou2* is required for early endodermal development (Reim and Brand, 2006), and mouse *Oct4* is expressed maternally in the oocyte (Ovitt and Scholer, 1998). However, the role in pluripotency of mouse *Oct4* is related to its zygotic expression, and the restricted expression of *Oct4* in the blastocyst is under the control of regulatory elements that are activated zygotically and not in the oocyte (Yeom *et al.*, 1996). Therefore, a role for maternal expression of these genes in pluripotency in the chick would imply a switch from maternal to zygotic control in the evolution of mammals, which is highly unlikely.

Taken together, the comparative expression analysis between pre-gastrulation mouse and chick embryos suggests that the EP-GRN was newly assembled in mammals, where expression of core transcription factors appeared in novel domains in space and time during evolution. Moreover microarray analysis complements and confirms the conclusions drawn from the in situ analysis, demonstrating that the pre-gastrulation chick embryo does not show a general enrichment in the expression of orthologues of

mouse pluripotency-related genes. This again suggests that the EP-GRN is not deployed in the same manner in the early mouse and chick embryo.

In line with the results presented here, recent work from our group has found differences in the expression patterns between chick and mouse for genes involved in the GRN that controls extraembryonic fate (Pernaute *et al.*, 2010). Although in the early chick embryo *Cdx2* and *Eomes* are expressed in extraembryonic tissues (area opaca), both differ from mouse in their time of expression, making it unlikely that the action of *Cdx2* upstream of *Eomes* seen in mouse also occurs in chick. In addition, unlike mouse, *Eomes* is expressed in the PGCs. This indicates that chick *Nanog* and *Pou2*-r cannot repress *Eomes* as occurs in mouse. What is more, orthologues of *Bmp4* and *Fgfr2*, which in mouse are expressed in TE and extraembryonic ectoderm, are not expressed in the early chick extraembryonic domains (Fig. 1). These findings suggest that not only was the EP-GRN newly assembled in mammalian evolution, but also that the appearance of the trophoectoderm was accompanied by critical changes in the GRN controlling the extraembryonic lineage.

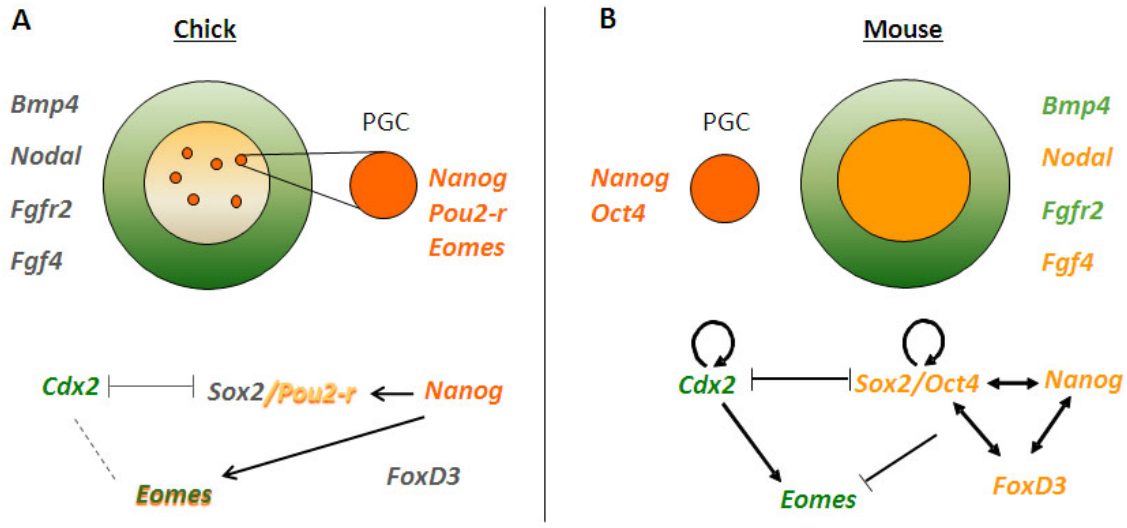


Figure 1. Diagram showing the differences between chick and mouse genes involved in embryonic pluripotency and specification of extraembryonic and PGC fates. During mammalian evolution, some genes whose orthologues were already present gained new expression domains and established new regulatory interactions. **(A)** Diagram of a EGK-X chick embryo, with the area opaca (outer circle, green), the area pellucida (inner circle, light orange) and PGCs scattered over the latter (small orange circles). While *Cdx2* and *Eomes* are expressed in the extraembryonic domain (area opaca), *Pou2*-r is expressed in the embryonic domain (area pellucida). This gene, together with *Eomes* and *Nanog*, is expressed in the PGCs. **(B)** Diagram of a cross-section through the mouse blastocyst perpendicular to the embryonic-abembryonic axis. While *Cdx2* and *Eomes* expression is conserved in the extraembryonic domain (green), two genes, *Bmp4* and *Fgfr2*, changed their expression in time and space and are now expressed in this domain. *Nanog* changed its spatial expression and is no longer restricted to the PGCs but is expressed in the whole embryonic domain (light orange). The paralogue of *Pou2*-r, *Oct4*, is also expressed in the embryonic domain together with *Sox2, Nodal*, *Fgf4* and *FoxD3*, which changed their expression in time and space. While *Oct4* is found in PGCs in analogy with its *Pou2-r* paralogue, *Nanog* expression is conserved and *Eomes* expression is lost. Genes in grey are not expressed in the chick embryo at this stage, black lines indicate evidence for interaction; grey lines indicate no evidence.

2. Recruiting genes to the EP-GRN

Genomic analysis shows that some EP-GRN genes are specific to mammals and in many instances arose by gene duplication, such as *Gdf3* (Canon *et al.*, 2006) and *Rex1* (Kim *et al.*, 2007). Gene duplication events generally provide a chance to evolve new gene functions, for example via the establishment of new protein-protein interactions (Wagner and Lynch, 2005) as suggested in the case of *Oct4* (Niwa *et al.*, 2008).

Therefore, it is tempting to speculate that gene duplications were major events behind the emergence of the pluripotency regulatory network. However, targets of core pluripotency factors (Marson *et al.*, 2008) are not enriched in mammalian specific genes. Furthermore, we found that the duplication of *Oct4* occurred before the divergence of amniotes, indicating that all core pluripotency factors were present long before the appearance of mammals. Therefore, if the EP-GRN was newly assembled in mammals, as the expression data suggest, it must have occurred not only by the appearance of novel genes but also through the recruitment of pre-existing ones (Fig. 2).

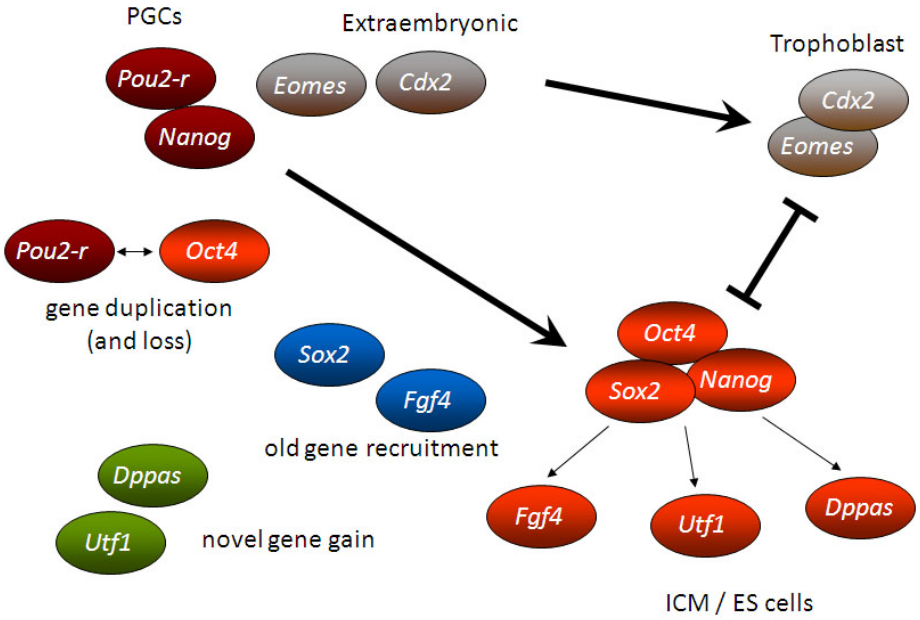


Figure 2. How the EP-GRN was built. EP-GRN arose by co-option (*Nanog*, *Sox2*, *Fgf4*), duplication (*Pou2*-r) and the appearance of novel genes (*Dppas*, *Utf1*), as well as new regulatory interactions that recruited new direct targets of the core factors or established reciprocal inhibition loops between the embryonic pluripotency core factors (*Nanog*, *Oct4* and *Sox2*) and extraembryonic specifiers (*Cdx2*, *Eomes*).

The most obvious way in which genes can gain new expression domains is through the appearance of novel cis-regulatory elements in their vicinity, allowing novel regulatory links (Wagner and Lynch, 2005). We find that the genomic regions bound by the core EP-GRN factors show very little conservation between mouse and chick, in accordance with them being new elements in the mammalian lineage. Further evidence for this comes from the fact that these regions are enriched in rodent- and mouse-specific

sequences, reflecting their recent evolutionary appearance. In contrast, rodent- and mouse-specific sequences are not found in the Gli3-bound regions of the more ancient limb-specification GRN (Vokes *et al.*, 2008). Within the mammalian lineage, the lower degree of conservation we found in regions bound by Oct4 and Nanog compared with the GRN controlling limb development (Results Fig. 11A) is in agreement with recent findings on the re-wiring of peripheral components of the EP-GRN (Kunarso *et al.*, 2010; Xie *et al.*, 2010).

A more informative analysis of genes targeted by Oct4 and Nanog could in principle be made by comparing genome-wide binding profiles for these genes in chick and mouse, such analysis would be extremely informative regarding the evolution of the network. However, this approach is currently not possible, because the antibodies to the mouse proteins do not cross-react with their chick homologues, precluding their use in chick ChIP-sequence approaches.

Our prediction that novel elements regulated by core pluripotency factors first appeared in mammals were confirmed by two examples. The sequences surrounding the Sox2/Oct4 cassette, which are critical for the expression of *Fgf4* and *Sox2* in the pre-implantation embryo and ES cells (Fraidenraich *et al.*, 1998; Tomioka *et al.*, 2002; Yuan *et al.*, 1995), are sufficiently conserved between mammals and chick to allow us to identify critical nucleotide changes that resulted in the appearance of adjacent functional HMG and POU binding sites. Functional assays in ES cells and pre-implantation embryos with the mouse enhancers and the corresponding chick genomic fragments, together with mutated versions where the sites were swapped between mouse and chick, show that the Sox2/Oct4 cassette is necessary but not sufficient for activity in pluripotent cells. This indicates that the mouse EP enhancers contain other non-conserved sequences necessary for activity that remain to be identified. Our results also demonstrate that the chick sequences, although similar enough to the mouse to identify the nucleotides corresponding to the Oct4/Sox2 cassette, do not respond to endogenous *Oct4* or *Sox2*. An obvious interpretation is that the nucleotide differences we observe in the chick correlate with changes in the binding specificities of the factors. Therefore, we investigated the potential of both mouse *Oct4* and chick *Pou2-r*, which show divergent protein sequences, to bind to pluripotency enhancers.

The results show that the proteins per se are largely interchangeable and are capable of acting equally on the mouse *Oct4* distal enhancer. This assay extends previous findings (Lavial *et al.*, 2007) that showed that *Pou2-r* was able to substitute the endogenous *Oct4* in mouse ES cells and activate the proximal *Oct4* promoter. However, the promoter used by these authors, while active in EpiS cells and in the epiblast, is not pluripotency-specific.

Our results suggest that the major changes that have occurred during evolution are mainly due to changes in the wiring of the network, and not to changes in the binding specificities of the factors. To further test this we also examined the ability of Oct4 and Pou2-r to activate the *Fgf4* and *Sox2* pluripotency enhancers, obtaining similar results to those for the *Oct4*-distal enhancer. More interestingly, we also found that chick *Pou2-r* cannot bind to the chick genomic regions homologous to these enhancers, and neither the mouse enhancers where the critical residues in the binding site for *Oct4* have been mutated to those present in the chick. This set of experiments therefore demonstrates that the changes in the mouse genome do not simply represent a change in binding site preference of the core factors.

3. What is special about embryonic pluripotency in mammals?

Our results raise the question of why a new gene regulatory network responsible for maintaining embryonic pluripotency should have arisen in mammals. We believe the answer may lie in the peculiar nourishment requirements of early mammalian embryos. In the embryos of most vertebrates (and invertebrates), the egg is supplied with yolk that feeds the embryo throughout development. This kind of embryo grows by a series of quick cell divisions (O'Farrell *et al.*, 2004), and patterning is initiated shortly after fertilization by maternal and/or external factors (such as the sperm entry point in amphibians, or gravitation in the chick). The situation is radically different in mammals, where eggs are devoid of yolk and embryo nourishment must be obtained through direct contact with the maternal uterine wall via the placenta. This change has been accompanied by the loss of genes encoding vitellogenin proteins, stored in the yolk of other species as a nutritional reservoir (Brawand *et al.*, 2008), and the appearance of a new cellular population, the trophoblast, which gives rise to the

majority of the placenta. In this regard, we also found striking differences in the expression of orthologues of extraembryonic genes in the early chick embryo (Pernaute *et al.*, 2010). Initial cell divisions of the mammalian embryo, as observed in mouse, are extremely slow (O'Farrell *et al.*, 2004), and until implantation there is hardly any growth. Recent evidence indicates that embryonic patterning in mammals is not initiated until peri-implantation stages, long after fertilization (Rossant and Tam, 2009). In fact, the first differentiation event to occur is the specification of the trophectoderm lineage, which is unique to mammals. Another key characteristic of the mouse embryo is that patterning is initiated by intrinsic and zygotic mechanisms: no external trigger or maternal signal is needed to establish the mayor body axes. This is most clearly illustrated in tetraploid aggregation chimeras, where nearly the whole embryo is derived from ES cells, which do not retain any spatial patterning information after prolonged growth and passage in tissues (Eakin and Hadjantonakis, 2006).

Therefore in mammalian embryos the delay in pattern formation and the uncoupling from external cues necessitate prolonged maintenance of embryonic cells in an undetermined and quiescent state (Fig. 3). The evolutionary appearance of the EP-GRN would have permitted this state, with the core factors (*Oct4-Sox2-Nanog*) silencing genes involved in early differentiation and specification (Boyer *et al.*, 2005; Loh *et al.*, 2006) while leaving them ready to be expressed in a controlled spatial and temporal fashion. Epigenetic modifiers might contribute to the dual situation of those genes. Recent studies have shown that a subset of developmental genes is modified by so-called bivalent domains. These domains are a combination of activating and repressive histone modifications, therefore contributing to silencing developmental genes while keeping them poised for activation (Bernstein *et al.*, 2006).

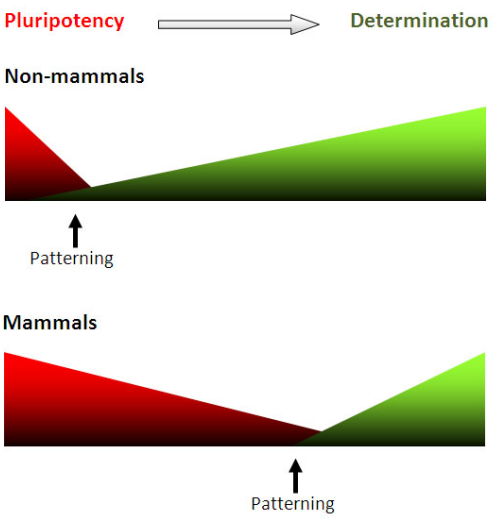


Figure 3. Differences between non-mammalian and mammalian early development. In non-mammals early patterning is driven by asymmetrically distributed maternal determinants or external cues; in mammals, in contrast, patterning is driven by zygotic transcription. This loss of dependence on maternal or external cues considerably delayed the onset of patterning in mammals. Therefore, mammalian embryos needed an extended period in which embryonic cells were maintained in a pluripotent state before differentiation.

Some light on the origin of the EP-GRN is shed by the fact that *Nanog* and *Oct4* homologues of both mouse and chick embryos are expressed in PGCs (Canon *et al.*, 2006; Lavial *et al.*, 2007). PGCs resemble ES cells in many respects (Seydoux and Braun, 2006; Zwaka and Thomson, 2005), and a possible scenario would be that *Nanog* and *Pou5* genes act in PGC to prevent their differentiation and maintain their germline potential until differentiation of the mature germ cells. The requirement for both factors in mouse germ cell development has been shown (Chambers *et al.*, 2007; Kehler *et al.*, 2004; Okamura *et al.*, 2008; Yamaguchi *et al.*, 2009) and pluripotency maintenance in PGCs may therefore be an older, evolutionarily conserved role for these genes in vertebrates that was later co-opted and expanded to form the network responsible for the same function in early embryonic cells of mammals.

Our results also have implications regarding the nature of chick ES cells, which might originate from PGCs in the early chick embryo. Chick ES cells have been derived from EGK-X blastoderm cells and maintained in an undifferentiated state in culture. These cells can contribute to somatic tissues in chimaeras (Petitte *et al.*, 2004) but they do not contribute (van de Lavoir *et al.*, 2006b) or only very poorly (Carsience *et al.*, 1993; Pain *et al.*, 1996) to the germ line. Another difference with mouse ES cells is that undifferentiated growth of chick cells requires Fgf in the culture medium (Lavial *et al.*, 2007), while this growth factor promotes differentiation of mouse ES cells (Silva and Smith, 2008). Cytokines and culture conditions for chick ES cells are highly similar to those used to amplify chick embryonic germ (EG) cells from gonadal cells or culture

PGCs *in vitro* (Lavial and Pain, 2009). Moreover, chick ES cells express low levels of the *Vasa* homologue *Cvh* (Lavial *et al.*, 2009), some germinal-associated genes and the chick homologue of *Eomes* (Lavial and Pain, 2009), whose expression is clear in chick PGCs (Pernaute *et al.*, 2010) but is absent in mouse ES cells.

These marked differences with mouse ES cells might be explained if chick blastoderm-derived pluripotent cells are derived from PGCs present in the freshly dissociated epiblast. Some of the few PGCs might gain a proliferative advantage and maintain an undifferentiated state as a result of the culture conditions, coming to resemble an EG cell type (van de Lavoir *et al.*, 2006a). Therefore, mouse ES cells and chick blastoderm-derived pluripotent cells might have a different embryonic origin. Our hypothesis questions the possibility of reprogramming somatic cells in avian species.

4. *Nanog* in early differentiation of neural and haematopoietic precursors

Nanog overexpression causes neural defects in mouse and chick embryos. Our results show altered expression levels of genes involved in neural differentiation, such as *Fgf8* and *Otx2*. Similarly to our Nanog overexpressing embryos, null mutants for either of these genes present defects in rostral brain development. Interestingly *Fgf8* and *Otx2* are among the earliest genes to be expressed in the epiblast and they are required for proper gastrulation (Ang *et al.*, 1996; Sun *et al.*, 1999). Given that *Nanog* transcripts are still expressed in the epiblast until early streak stages (E6.5), *Fgf8* and *Otx2* could be *Nanog* targets in the early epiblast. This is supported by the fact that Nanog binding sites were found by ChIP in the vicinity of both genes (Chen *et al.*, 2008; Marson *et al.*, 2008). Therefore *Nanog* might repress *Fgf8* and *Otx2* at early gastrulation, and the Nanog overexpression we induce at later stages would recapitulate the regulatory interactions that occur in the early epiblast.

We found that *Fgf8* expression was exclusively altered in the isthmus. One possibility is that the developmental state or history of this territory confers on it a different potential to respond to the *Fgf8* inductive signal, a competence other *Fgf8* expression domains do not have. Interestingly, mutants for zebrafish *pou2*/spg, which is expressed in the neural plate, are characterized by absence of the isthmus, altered *Fgf8* expression in this region, and changes in forebrain markers (Reim and Brand,

2002). This is very similar to what we observed in Nanog overexpressing embryos. The authors demonstrate that the competence to respond to *Fgf8* signalling during the establishment and maintenance of the mid-hindbrain is mediated by *pou2/spg,* a function which might be conserved in other vertebrates. In chick, functional conservation might be indicated by the strong expression of *Pou2*-r in the mid-hindbrain from HH8 (Lavial *et al.*, 2007). However, *Oct4* is not expressed in the neural plate in mouse.

Chick *Nanog* is also expressed in the anterior neural plate from HH6 (Lavial *et al.*, 2007). However, *Nanog* is not expressed either in the neural plate in mouse. Therefore, in mammals, even though *Nanog*'s regulatory inputs into neural differentiation might be conserved–as suggested by Nanog overexpression results, its neural expression is lost.

Mouse embryos overexpressing Nanog also present alterations in genes related to haematological development. Similar to the situation with neural genes, *Nanog* might repress those genes involved in haematopoietic differentiation at the gastrulation stage.

Therefore, *Nanog*, in addition to controlling pluripotency, might be important at gastrulation stages, repressing genes involved in the differentiation of certain precursors. Further analysis must be conducted to confirm this function and study its conservation during evolution.

Our research has extended our understanding of mammalian embryonic pluripotency, providing a more complete picture of its control and evolution, which is an important step in this exciting era of the regenerative medicine.

CONCLUSIONS

1. The EP-GRN was newly assembled in mammals. Orthologues of many genes involved in embryonic pluripotency and early lineages in mouse are not expressed in equivalent territories of the early chick embryo. These chick embryos and their derived cells in culture differ in their global expression profiles from the mouse pluripotent state.

2. The EP-GRN was partly assembled through the appearance of new pluripotent genes that in many instances arose by duplication. However, such events were not the main driver of the emergence of the EP-GRN. First, all core pluripotency factors were present long before the appearance of mammals. Second, downstream targets of the mouse core pluripotency factors are not enriched in mammalian specific genes.

3. The EP-GRN was mainly assembled by co-opting pre-existing genes through the appearance of core EP factor response elements. The genomic sequence elements which mediate the downstream action of the mouse core pluripotency factors are non-conserved with chick.

4. The major changes in the appearance of the EP-GRN were in the wiring of the network. The new wiring by novel response elements allowed the core factors to gain new expression domains and/or regulatory interactions. Interestingly these core factors did not change their binding specificities during the evolution.

5. The novel regulatory elements appeared by modest changes such as single nucleotide mutations or insertions. These types of changes created novel Oct4/Sox2 binding sites in the mouse *Fgf4* and *Sox2* pluripotency genes. The corresponding sequences in chick are unable to bind *Oct4* and *Sox2*.

6. The EP-GRN might have arisen to support the peculiar nourishment requirements of early mammalian embryos, which necessitate prolonged maintenance of embryonic cells in an undetermined state before engaging in differentiation programs.

7. *Nanog* might have conserved functions additional to controlling pluripotency, acting as a repressor of neural and haematopoietic differentiation at gastrulation stages.

1. El ensamblaje de la red génica que regula la pluripotencia embrionaria ocurrió en la evolución durante la aparición de los mamíferos. Hemos encontrado que los ortólogos de genes involucrados en la pluripotencia embrionaria y en los linajes tempranos de ratón, no se expresan en el embrión temprano de pollo en los territorios considerados equivalentes al embrión de ratón. Además, el perfil de expresión global del estadio pluripotente de ratón es muy diferente a los perfiles que encontramos en los embriones tempranos de pollo y a las células derivadas de éstos.

2. La aparición de nuevos genes de pluripotencia que en muchos casos surgieron por duplicaciones, fue importante para el ensamblaje de la red génica regulatoria de la pluripotencia embrionaria. Sin embargo estos eventos no fueron los grandes impulsores de la aparición de la red de pluripotencia embrionaria. Primero porque todos los factores centrales de la pluripotencia ya estaban presentes antes de la aparición de los mamíferos. Segundo porque hemos visto que en el ratón entre las dianas *downstream* a los factores centrales de pluripotencia no abundan los genes exclusivos de mamíferos.

3. La red génica regulatoria de la pluripotencia embrionaria se ensambló principalmente cooptando genes ya existentes, mediante la aparición de elementos de respuesta a los factores centrales de la pluripotencia embrionaria. De hecho hemos encontrado que no están conservados en pollo los elementos de secuencia génomica que median la acción downstream a los factores centrales de pluripotencia en ratón.

4. Los cambios que tuvieron lugar en las conexiones fueron los que impulsaron la aparición de la red génica de la pluripotencia embrionaria. Las nuevas conexiones creadas por los nuevos elementos de respuesta permitieron a los factores centrales adquirir nuevos dominios de expresión y nuevas interacciones regulatorias. Sin embargo los factores centrales no cambiaron sus especificidades de unión a lo largo de la evolución.

5. Los nuevos elementos reguladores surgieron por pequeños cambios, tales como mutaciones de un solo nucleótido o inserciones. Este tipo de cambios fueron probablemente los que crearon los nuevos sitios de unión para *Oct4* y *Sox2* en los genes de pluripotencia *Fgf4* y *Sox2*. De hecho encontramos que en pollo las secuencias correspondientes a los sitios de unión son incapaces de unir los dos factores.

6. La aparición en mamíferos de una red de regulación de la pluripotencia embrionaria quizá no sea casual sino relacionada con las peculiaridades y requerimientos nutricionales del embrión temprano de mamíferos. Éste necesitaba prolongar el mantenimiento de las células embrionarias en un estado indeterminado antes de embarcarse en procesos de diferenciación y la nueva red se lo habría permitido.

7. *Nanog* podría, aparte de regular la pluripotencia, tener un papel adicional y conservado a lo largo de la evolución como represor de la diferenciación neural y hematopoyética en los estadios de gastrulación.

BIBLIOGRAPHY

Acloque H, Wilkinson DG, Nieto MA (2008). In situ hybridization analysis of chick embryos in whole-mount and tissue sections. *Methods Cell Biol* **87:** 169-85.

Andras Nagy MG, Kristina Vintersten, Richard Behringer (2003). Manipulating the mouse embryo: A laboratory manual *Cold Spring Harbor, New York: Cold Spring Habor laboratory*.

Ang SL, Jin O, Rhinn M, Daigle N, Stevenson L, Rossant J (1996). A targeted mouse Otx2 mutation leads to severe defects in gastrulation and formation of axial mesoderm and to deletion of rostral brain. *Development* **122:** 243-52.

Ariza-McNaughton L, Krumlauf R (2002). Non-radioactive in situ hybridization: simplified procedures for use in whole-mounts of mouse and chick embryos. *Int Rev Neurobiol* **47:** 239-50.

Arman E, Haffner-Krausz R, Chen Y, Heath JK, Lonai P (1998). Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. *Proc Natl Acad Sci U S A* **95:** 5082-7.

Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17:** 126-40.

Bachvarova RF, Masi T, Drum M, Parker N, Mason K, Patient R *et al* (2004). Gene expression in the axolotl germ line: Axdazl, Axvh, Axoct-4, and Axkit. *Dev Dyn* **231:** 871-80.

Beck F, Erler T, Russell A, James R (1995). Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. *Dev Dyn* **204:** 219-27.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J *et al* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125:** 315-26.

Boardman PE, Sanz-Ezquerro J, Overton IM, Burt DW, Bosch E, Fong WT *et al* (2002). A comprehensive collection of chicken cDNAs. *Curr Biol* **12:** 1965-9.

Boiani M, Scholer HR (2005). Regulatory networks in embryo-derived pluripotent stem cells. *Nat Rev Mol Cell Biol* **6:** 872-84.

Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19:** 185-93.

Bortvin A, Eggan K, Skaletsky H, Akutsu H, Berry DL, Yanagimachi R *et al* (2003). Incomplete reactivation of Oct4-related genes in mouse embryos cloned from somatic nuclei. *Development* **130:** 1673-80.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP *et al* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122:** 947-56.

Brawand D, Wahli W, Kaessmann H (2008). Loss of egg yolk genes in mammals and the origin of lactation and placentation. *PLoS Biol* **6:** e63.

Brennan J, Lu CC, Norris DP, Rodriguez TA, Beddington RS, Robertson EJ (2001). Nodal signalling in the epiblast patterns the early mouse embryo. *Nature* **411:** 965-9.

Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM *et al* (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448:** 191-5.

Burgess S, Reim G, Chen W, Hopkins N, Brand M (2002). The zebrafish spiel-ohne-grenzen (spg) gene encodes the POU domain protein Pou2 related to mammalian Oct4 and is essential for formation of the midbrain and hindbrain, and for pre-gastrula morphogenesis. *Development* **129:** 905-16.

Camp EM, Sanchez-Sanchez AV, Garcia-Espana A, Desalle R, Odqvist L, O'Connor JE *et al* (2009). Nanog regulates proliferation during early fish development. *Stem Cells*.

Canon S, Herranz C, Manzanares M (2006). Germ cell restricted expression of chick Nanog. *Dev Dyn* **235:** 2889-94.

Carsience RS, Clark ME, Verrinder Gibbins AM, Etches RJ (1993). Germline chimeric chickens from dispersed donor blastodermal cells and compromised recipient embryos. *Development* **117:** 669-75.

Ciruna BG, Rossant J (1999). Expression of the T-box gene Eomesodermin during early mouse development. *Mech Dev* **81:** 199-203.

Clark AT, Rodriguez RT, Bodnar MS, Abeyta MJ, Cedars MI, Turek PJ *et al* (2004). Human STELLAR, NANOG, and GDF3 genes are expressed in pluripotent cells and map to chromosome 12p13, a hotspot for teratocarcinoma. *Stem Cells* **22:** 169-79.

Cockburn K, Rossant J (2010). Making the blastocyst: lessons from the mouse. *J Clin Invest* **120:** 995-1003.

consortium ICGs (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432:** 695-716.

Crossley PH, Martinez S, Martin GR (1996). Midbrain development induced by FGF8 in the chick embryo. *Nature* **380:** 66-8.

Chai N, Patel Y, Jacobson K, McMahon J, McMahon A, Rappolee DA (1998). FGF is an essential regulator of the fifth cell division in preimplantation mouse embryos. *Dev Biol* **198:** 105-15.

Chakravarthy H, Boer B, Desler M, Mallanna SK, McKeithan TW, Rizzino A (2008). Identification of DPPA4 and other genes as putative Sox2:Oct-3/4 target genes using a combination of in silico analysis and transcription-based assays. *J Cell Physiol* **216:** 651-62.

Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S *et al* (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113:** 643-55.

Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, Robertson M *et al* (2007). Nanog safeguards pluripotency and mediates germline development. *Nature* **450:** 1230-4.

Chambers I, Tomlinson SR (2009). The transcriptional foundation of pluripotency. *Development* **136:** 2311-22.

Chapman SC, Schubert FR, Schoenwolf GC, Lumsden A (2002). Analysis of spatial and temporal gene expression patterns in blastula and gastrula stage chick embryos. *Dev Biol* **245:** 187-99.

Chen L, Yabuuchi A, Eminli S, Takeuchi A, Lu CW, Hochedlinger K *et al* (2009). Cross-regulation of the Nanog and Cdx2 promoters. *Cell Res* **19:** 1052-61.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB *et al* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133:** 1106-17.

Chew JL, Loh YH, Zhang W, Chen X, Tam WL, Yeap LS *et al* (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol* **25:** 6031-46.

Davidson E, Levin M (2005). Gene regulatory networks. *Proc Natl Acad Sci U S A* **102:** 4935.

Deakin JE, Papenfuss AT, Belov K, Cross JG, Coggill P, Palmer S *et al* (2006). Evolution and comparative analysis of the MHC Class III inflammatory region. *BMC Genomics* **7:** 281.

Dietrich JE, Hiiragi T (2007). Stochastic patterning in the mouse pre-implantation embryo. *Development* **134:** 4219-31.

Dixon JE, Allegrucci C, Redwood C, Kump K, Bian Y, Chatfield J *et al* (2010). Axolotl Nanog activity in mouse embryonic stem cells demonstrates that ground state pluripotency is conserved from urodele amphibians to mammals. *Development* **137:** 2973-80.

Eakin GS, Hadjantonakis AK (2006). Production of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. *Nat Protoc* **1:** 1145-53.

Eyal-Giladi H, Kochav S (1976). From cleavage to primitive streak formation: a complementary normal table and a new look at the first stages of the development of the chick. I. General morphology. *Dev Biol* **49:** 321-37.

ENCODE Project consortium the human genome (2007). Identification and analysis of functional elements in 1% of by the ENCODE pilot project. *Nature* **447:** 799-816.

Feldman B, Poueymirou W, Papaioannou VE, DeChiara TM, Goldfarb M (1995). Requirement of FGF-4 for postimplantation mouse development. *Science* **267:** 246-9.

Felsenstein J (1996). Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol* **266:** 418-27.

Fraidenraich D, Lang R, Basilico C (1998). Distinct regulatory elements govern Fgf4 gene expression in the mouse blastocyst, myotomes, and developing limb. *Dev Biol* **204:** 197-209.

Frankenberg S, Pask A, Renfree MB (2010). The evolution of class V POU domain transcription factors in vertebrates and their characterisation in a marsupial. *Dev Biol* **337:** 162-70.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res* **32:** W273-9.

Ginis I, Luo Y, Miura T, Thies S, Brandenberger R, Gerecht-Nir S *et al* (2004). Differences between human and mouse embryonic stem cells. *Dev Biol* **269:** 360-80.

Graves JA, Westerman M (2002). Marsupial genetics and genomics. *Trends Genet* **18:** 517-21.

Guzman-Ayala M, Ben-Haim N, Beck S, Constam DB (2004). Nodal protein processing and fibroblast growth factor 4 synergize to maintain a trophoblast stem cell microenvironment. *Proc Natl Acad Sci U S A* **101:** 15656-60.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41:** 95-98.

Hamburger V, Hamilton HL (1992). A series of normal stages in the development of the chick embryo. 1951. *Dev Dyn* **195:** 231-72.

Hanna LA, Foreman RK, Tarasenko IA, Kessler DS, Labosky PA (2002). Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. *Genes Dev* **16:** 2650-61.

Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP *et al* (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432:** 695-716.

Hochedlinger K, Yamada Y, Beard C, Jaenisch R (2005). Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* **121:** 465-77.

Houbaviy HB, Dennis L, Jaenisch R, Sharp PA (2005). Characterization of a highly variable eutherian microRNA gene. *Rna* **11:** 1245-57.

Houbaviy HB, Murray MF, Sharp PA (2003). Embryonic stem cell-specific MicroRNAs. *Dev Cell* **5:** 351-8.

Hughes AL, Friedman R (2008). Genome size reduction in the chicken has involved massive loss of ancestral protein-coding genes. *Mol Biol Evol* **25:** 2681-8.

Inoue N, Hess KD, Moreadith RW, Richardson LL, Handel MA, Watson ML *et al* (1999). New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. *Hum Mol Genet* **8:** 1201-7.

Judson RL, Babiarz JE, Venere M, Blelloch R (2009). Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat Biotechnol* **27:** 459-61.

Kalmar T, Lim C, Hayward P, Munoz-Descalzo S, Nichols J, Garcia-Ojalvo J *et al* (2009). Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol* **7:** e1000149.

Kaufmann (1992). *The atlas of mouse development*. Elsevier Academic Press.

Kehler J, Tolkunova E, Koschorz B, Pesce M, Gentile L, Boiani M *et al* (2004). Oct4 is required for primordial germ cell survival. *EMBO Rep* **5:** 1078-83.

Kim J, Chu J, Shen X, Wang J, Orkin SH (2008). An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132:** 1049-61.

Kim JD, Faulk C, Kim J (2007). Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1. *Nucleic Acids Res* **35:** 3442-52.

Kim SK, Suh MR, Yoon HS, Lee JB, Oh SK, Moon SY *et al* (2005). Identification of developmental pluripotency associated 5 expression in human pluripotent stem cells. *Stem Cells* **23:** 458-62.

Kos R, Reedy MV, Johnson RL, Erickson CA (2001). The winged-helix transcription factor FoxD3 is important for establishing the neural crest lineage and repressing melanogenesis in avian embryos. *Development* **128:** 1467-79.

Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS *et al* (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* **42:** 631-4.

Kunath T, Strumpf D, Rossant J (2004). Early trophoblast determination and stem cell maintenance in the mouse--a review. *Placenta* **25 Suppl A:** S32-8.

Kuntz S, Kieffer E, Bianchetti L, Lamoureux N, Fuhrmann G, Viville S (2008). Tex19, a mammalian-specific protein with a restricted expression in pluripotent stem cells and germ line. *Stem Cells* **26:** 734-44.

Kuroda T, Tada M, Kubota H, Kimura H, Hatano SY, Suemori H *et al* (2005). Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. *Mol Cell Biol* **25:** 2475-85.

Lavial F, Acloque H, Bachelard E, Nieto MA, Samarut J, Pain B (2009). Ectopic expression of Cvh (Chicken Vasa homologue) mediates the reprogramming of chicken embryonic stem cells to a germ cell fate. *Dev Biol* **330:** 73-82.

Lavial F, Acloque H, Bertocchini F, Macleod DJ, Boast S, Bachelard E *et al* (2007). The Oct4 homologue PouV and Nanog regulate pluripotency in chicken embryonic stem cells. *Development* **134:** 3549-63.

Lavial F, Pain B (2009). Chicken embryonic stem cells as a non-mammalian embryonic stem cell model. *Dev Growth Differ* **52:** 101-14.

Leitch HG, Blair K, Mansfield W, Ayetey H, Humphreys P, Nichols J *et al* Embryonic germ cells from mice and rats exhibit properties consistent with a generic pluripotent ground state. *Development* **137:** 2279-87.

Li JY, Joyner AL (2001). Otx2 and Gbx2 are required for refinement and not induction of mid-hindbrain gene expression. *Development* **128:** 4979-91.

Liu A, Losos K, Joyner AL (1999). FGF8 can activate Gbx2 and transform regions of the rostral mouse brain into a hindbrain fate. *Development* **126:** 4827-38.

Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X *et al* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38:** 431-40.

Luo J, Sladek R, Bader JA, Matthyssen A, Rossant J, Giguere V (1997). Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. *Nature* **388:** 778-82.

Macarthur BD, Ma'ayan A, Lemischka IR (2009). Systems biology of stem cell fate and cellular reprogramming. *Nat Rev Mol Cell Biol* **10:** 672-81.

Maldonado-Saldivia J, van den Bergen J, Krouskos M, Gilchrist M, Lee C, Li R *et al* (2007). Dppa2 and Dppa4 are closely linked SAP motif genes restricted to pluripotent cells and the germ line. *Stem Cells* **25:** 19-28.

Marom K, Shapira E, Fainsod A (1997). The chicken caudal genes establish an anterior-posterior gradient by partially overlapping temporal and spatial patterns of expression. *Mech Dev* **64:** 41-52.

Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S *et al* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134:** 521-33.

Martinez-Barbera JP, Signore M, Boyl PP, Puelles E, Acampora D, Gogoi R *et al* (2001). Regionalisation of anterior neuroectoderm and its competence in responding to forebrain and midbrain inducing activities depend on mutual antagonism between OTX2 and GBX2. *Development* **128:** 4789-800.

Masaki H, Nishida T, Kitajima S, Asahina K, Teraoka H (2007). Developmental pluripotency-associated 4 (DPPA4) localized in active chromatin inhibits mouse embryonic stem cell differentiation into a primitive ectoderm lineage. *J Biol Chem* **282:** 33034-42.

Mate KE, Robinson ES, Vandeberg JL, Pedersen RA (1994). Timetable of in vivo embryonic development in the grey short-tailed opossum (Monodelphis domestica). *Mol Reprod Dev* **39:** 365-74.

Messerschmidt DM, Kemler R (2010). Nanog is required for primitive endoderm formation through a non-cell autonomous mechanism. *Dev Biol* **344:** 129-37.

Miyagi S, Masui S, Niwa H, Saito T, Shimazaki T, Okano H *et al* (2008). Consequence of the loss of Sox2 in the developing brain of the mouse. *FEBS Lett* **582:** 2811-5.

Monk M, Hitchins M, Hawes S (2008). Differential expression of the embryo/cancer gene ECSA(DPPA2), the cancer/testis gene BORIS and the pluripotency structural gene OCT4, in human preimplantation development. *Mol Hum Reprod* **14:** 347-55.

Morrison GM, Brickman JM (2006). Conserved roles for Oct4 homologues in maintaining multipotency during early vertebrate development. *Development* **133:** 2011-22.

Mullin NP, Yates A, Rowe AJ, Nijmeijer B, Colby D, Barlow PN *et al* (2008). The pluripotency rheostat Nanog functions as a dimer. *Biochem J* **411:** 227-31.

Nagy A, Gertsensten M, Vintersten K, Behringer R (2003). Manipulating the mouse embryo: a laboratory manual, 3rd edn. *Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York.*

Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T *et al* (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* **26:** 101-6.

Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I *et al* (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95:** 379-91.

Nishimoto M, Miyagi S, Yamagishi T, Sakaguchi T, Niwa H, Muramatsu M *et al* (2005). Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus. *Mol Cell Biol* **25:** 5084-94.

Niswander L, Martin GR (1992). Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. *Development* **114:** 755-68.

Niwa H (2007). How is pluripotency determined and maintained? *Development* **134:** 635-46.

Niwa H, Miyazaki J, Smith AG (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24:** 372-6.

Niwa H, Sekita Y, Tsend-Ayush E, Grutzner F (2008). Platypus Pou5f1 reveals the first steps in the evolution of trophectoderm differentiation and pluripotency in mammals. *Evol Dev* **10:** 671-82.

Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, Yagi R *et al* (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123:** 917-29.

Niwa H, Yamamura K, Miyazaki J (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108:** 193-9.

O'Farrell PH, Stumpff J, Su TT (2004). Embryonic cleavage cycles: how is a mouse like a fly? *Curr Biol* **14:** R35-45.

Odom DT, Dowell RD, Jacobsen ES, Gordon W, Danford TW, MacIsaac KD *et al* (2007). Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat Genet* **39:** 730-2.

Okamura D, Tokitake Y, Niwa H, Matsui Y (2008). Requirement of Oct3/4 function for germ cell specification. *Dev Biol* **317:** 576-84.

Okuda A, Fukushima A, Nishimoto M, Orimo A, Yamagishi T, Nabeshima Y *et al* (1998). UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. *Embo J* **17:** 2019-32.

Onichtchouk D, Geier F, Polok B, Messerschmidt DM, Mossner R, Wendik B *et al* Zebrafish Pou5f1-dependent transcriptional networks in temporal control of early development. *Mol Syst Biol* **6:** 354.

Ovitt CE, Scholer HR (1998). The molecular biology of Oct-4 in the early mouse embryo. *Mol Hum Reprod* **4:** 1021-31.

Page RD (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12:** 357-8.

Pain B, Clark ME, Shen M, Nakazawa H, Sakurai M, Samarut J *et al* (1996). Long-term in vitro culture and characterisation of avian embryonic stem cells with multiple morphogenetic potentialities. *Development* **122:** 2339-48.

Pan G, Li J, Zhou Y, Zheng H, Pei D (2006). A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. *FASEB J* **20:** 1730-2.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M *et al* (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444:** 499-502.

Pennacchio LA, Loots GG, Nobrega MA, Ovcharenko I (2007). Predicting tissue-specific enhancers in the human genome. *Genome Res* **17:** 201-11.

Pera MF, Trounson AO (2004). Human embryonic stem cells: prospects for development. *Development* **131:** 5515-25.

Pernaute B, Canon S, Crespo M, Fernandez-Tresguerres B, Rayon T, Manzanares M (2010). Comparison of extraembryonic expression of Eomes and Cdx2 in pregastrulation chick and mouse embryo unveils regulatory changes along evolution. *Dev Dyn* **239:** 620-9.

Petitte JN, Liu G, Yang Z (2004). Avian pluripotent stem cells. *Mech Dev* **121:** 1159-68.

Pierre A, Gautier M, Callebaut I, Bontoux M, Jeanpierre E, Pontarotti P *et al* (2007). Atypical structure and phylogenomic evolution of the new eutherian oocyte- and embryo-expressed KHDC1/DPPA5/ECAT1/OOEP gene family. *Genomics* **90:** 583-94.

Popova NV, Morris RJ (2004). Genetic regulation of mouse stem cells: identification of two keratinocyte stem cell regulatory loci. *Curr Top Microbiol Immunol* **280:** 111-37.

Prabhakar S, Poulin F, Shoukry M, Afzal V, Rubin EM, Couronne O *et al* (2006). Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res* **16:** 855-63.

Rappolee DA, Basilico C, Patel Y, Werb Z (1994). Expression and function of FGF-4 in peri-implantation development in mouse embryos. *Development* **120:** 2259-69.

Reim G, Brand M (2002). Spiel-ohne-grenzen/pou2 mediates regional competence to respond to Fgf8 during zebrafish early neural development. *Development* **129:** 917-33.

Reim G, Brand M (2006). Maternal control of vertebrate dorsoventral axis formation and epiboly by the POU domain protein Spg/Pou2/Oct4. *Development* **133:** 2757-70.

Remenyi A, Lins K, Nissen LJ, Reinbold R, Scholer HR, Wilmanns M (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev* **17:** 2048-59.

Rex M, Orme A, Uwanogho D, Tointon K, Wigmore PM, Sharpe PT *et al* (1997). Dynamic expression of chicken Sox2 and Sox3 genes in ectoderm induced to form neural tissue. *Dev Dyn* **209:** 323-32.

Robertson M, Stenhouse F, Colby D, Marland JR, Nichols J, Tweedie S *et al* (2006). Nanog retrotransposed genes with functionally conserved open reading frames. *Mamm Genome* **17:** 732-43.

Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH *et al* (2005). Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem* **280:** 24731-7.

Rossant J (2004). Lineage development and polar asymmetries in the peri-implantation mouse blastocyst. *Semin Cell Dev Biol* **15:** 573-81.

Rossant J, Cross JC (2001). Placental development: lessons from mouse mutants. *Nat Rev Genet* **2:** 538-48.

Rossant J, Tam PP (2009). Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development* **136:** 701-13.

Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4:** 406-25.

Sanchez-Sanchez AV, Camp E, Garcia-Espana A, Leal-Tassias A, Mullor JL (2010). Medaka Oct4 is expressed during early embryo development, and in primordial germ cells and adult gonads. *Dev Dyn* **239:** 672-9.

Sato T, Joyner AL (2009). The duration of Fgf8 isthmic organizer expression is key to patterning different tectal-isthmo-cerebellum structures. *Development* **136:** 3617-26.

Sauka-Spengler T, Barembaum M (2008). Gain- and loss-of-function approaches in the chick embryo. *Methods Cell Biol* **87:** 237-56.

Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V *et al* (2008). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*.

Schepers GE, Teasdale RD, Koopman P (2002). Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. *Dev Cell* **3:** 167-70.

Scholer HR, Ruppert S, Suzuki N, Chowdhury K, Gruss P (1990). New type of POU domain in germ line-specific protein Oct-4. *Nature* **344:** 435-9.

Selwood L, Johnson MH (2006). Trophoblast and hypoblast in the monotreme, marsupial and eutherian mammal: evolution and origins. *Bioessays* **28:** 128-45.

Seydoux G, Braun RE (2006). Pathway to totipotency: lessons from germ cells. *Cell* **127:** 891-904.

Shamim H, Mason I (1999). Expression of Fgf4 during early development of the chick embryo. *Mech Dev* **85:** 189-92.

Sharov AA, Masui S, Sharova LV, Piao Y, Aiba K, Matoba R *et al* (2008). Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. *BMC Genomics* **9:** 269.

Silva J, Nichols J, Theunissen TW, Guo G, van Oosten AL, Barrandon O *et al* (2009). Nanog is the gateway to the pluripotent ground state. *Cell* **138:** 722-37.

Silva J, Smith A (2008). Capturing pluripotency. *Cell* **132:** 532-6.

Smith A (2005). The battlefield of pluripotency. *Cell* **123:** 757-60.

Smyth GK (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3:** Article3.

Soudais C, Bielinska M, Heikinheimo M, MacArthur CA, Narita N, Saffitz JE *et al* (1995). Targeted mutagenesis of the transcription factor GATA-4 gene in mouse embryonic stem cells disrupts visceral endoderm differentiation in vitro. *Development* **121:** 3877-88.

Stern CD (2005). The chick; a great model system becomes even greater. *Dev Cell* **8:** 9-17.

Strumpf D, Mao CA, Yamanaka Y, Ralston A, Chawengsaksophak K, Beck F *et al* (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132:** 2093-102.

Stuhmer T, Anderson SA, Ekker M, Rubenstein JL (2002). Ectopic expression of the Dlx genes induces glutamic acid decarboxylase and Dlx expression. *Development* **129:** 245-52.

Sun X, Meyers EN, Lewandoski M, Martin GR (1999). Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes Dev* **13:** 1834-46.

Takahashi K, Yamanaka S (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126:** 663-76.

Takaoka K, Yamamoto M, Shiratori H, Meno C, Rossant J, Saijoh Y *et al* (2006). The mouse embryo autonomously acquires anterior-posterior polarity at implantation. *Dev Cell* **10:** 451-9.

Tallack MR, Whitington T, Yuen WS, Wainwright EN, Keys JR, Gardiner BB *et al* A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Res* **20:** 1052-63.

Tam PP, Rossant J (2003). Mouse embryonic chimeras: tools for studying mammalian development. *Development* **130:** 6155-63.

Tanaka TS, Lopez de Silanes I, Sharova LV, Akutsu H, Yoshikawa T, Amano H *et al* (2006). Esg1, expressed exclusively in preimplantation embryos, germline, and embryonic stem cells, is a putative RNA-binding protein with broad RNA targets. *Dev Growth Differ* **48:** 381-90.

Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X *et al* (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* **6:** 468-78.

Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL *et al* (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448:** 196-9.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25:** 4876-82.

Tomioka M, Nishimoto M, Miyagi S, Katayanagi T, Fukui N, Niwa H *et al* (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Res* **30:** 3202-13.

Tomotsune D, Shirai M, Takihara Y, Shimada K (2000). Regulation of Hoxb3 expression in the hindbrain and pharyngeal arches by rae28, a member of the mammalian Polycomb group of genes. *Mech Dev* **98:** 165-9.

Torres J, Watt FM (2008). Nanog maintains pluripotency of mouse embryonic stem cells by inhibiting NFkappaB and cooperating with Stat3. *Nat Cell Biol* **10:** 194-201.

Tsunekawa N, Naito M, Sakai Y, Nishida T, Noce T (2000). Isolation of chicken vasa homolog gene and tracing the origin of primordial germ cells. *Development* **127:** 2741-50.

Uchikawa M (2008). Enhancer analysis by chicken embryo electroporation with aid of genome comparison. *Dev Growth Differ* **50:** 467-74.

Uchikawa M, Ishida Y, Takemoto T, Kamachi Y, Kondoh H (2003). Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. *Dev Cell* **4:** 509-19.

Uchikawa M, Takemoto T, Kamachi Y, Kondoh H (2004). Efficient identification of regulatory sequences in the chicken genome by a powerful combination of embryo electroporation and genome comparison. *Mech Dev* **121:** 1145-58.

van de Lavoir MC, Diamond JH, Leighton PA, Mather-Love C, Heyer BS, Bradshaw R *et al* (2006a). Germline transmission of genetically modified primordial germ cells. *Nature* **441:** 766-9.

van de Lavoir MC, Mather-Love C, Leighton P, Diamond JH, Heyer BS, Roberts R *et al* (2006b). High-grade transgenic somatic chimeras from chicken embryonic stem cells. *Mech Dev* **123:** 31-41.

van den Boom V, Kooistra SM, Boesjes M, Geverts B, Houtsmuller AB, Monzen K *et al* (2007). UTF1 is a chromatin-associated protein involved in ES cell differentiation. *J Cell Biol* **178:** 913-24.

Vokes SA, Ji H, Wong WH, McMahon AP (2008). A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. *Genes Dev* **22:** 2651-63.

Wagner GP, Lynch VJ (2005). Molecular evolution of evolutionary novelties: the vagina and uterus of therian mammals. *J Exp Zool B Mol Dev Evol* **304:** 580-92.

Wang J, Levasseur DN, Orkin SH (2008). Requirement of Nanog dimerization for stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A* **105:** 6326-31.

Wang J, Rao S, Chu J, Shen X, Levasseur DN, Theunissen TW *et al* (2006). A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444:** 364-8.

Wilson SI, Rydstrom A, Trimborn T, Willert K, Nusse R, Jessell TM *et al* (2001). The status of Wnt signalling regulates neural and epidermal fates in the chick embryo. *Nature* **411:** 325-30.

Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP *et al* (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336:** 684-7.

Wolpert L, Beddington, R. S., Jessell, T. M., Lawrence, P., Meyerowitz, E., and Smith, J. (2002). Principles of developmen*t*, second edn. *Oxford university press: Oxford.*

Xie D, Chen CC, Ptaszek LM, Xiao S, Cao X, Fang F *et al* (2010). Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Res* **20:** 804-15.

Yamaguchi S, Kimura H, Tada M, Nakatsuji N, Tada T (2005). Nanog expression in mouse germ cell development. *Gene Expr Patterns* **5:** 639-46.

Yamaguchi S, Kurimoto K, Yabuta Y, Sasaki H, Nakatsuji N, Saitou M *et al* (2009). Conditional knockdown of Nanog induces apoptotic cell death in mouse migrating primordial germ cells. *Development* **136:** 4011-20.

Yamanaka Y, Ralston A, Stephenson RO, Rossant J (2006). Cell and molecular regulation of the mouse blastocyst. *Dev Dyn* **235:** 2301-14.

Yee SP, Rigby PW (1993). The regulation of myogenin gene expression during the embryonic development of the mouse. *Genes Dev* **7:** 1277-89.

Yeom YI, Fuhrmann G, Ovitt CE, Brehm A, Ohbo K, Gross M *et al* (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122:** 881-94.

Ying QL, Nichols J, Chambers I, Smith A (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115:** 281-92.

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S *et al* (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318:** 1917-20.

Yuan H, Corbi N, Basilico C, Dailey L (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* **9:** 2635-45.

Zagris N, Kalantzis K, Guialis A (1998). Activation of embryonic genome in chick. *Zygote* **6:** 227-31.

Zhao Y, Yin X, Qin H, Zhu F, Liu H, Yang W *et al* (2008). Two supporting factors greatly improve the efficiency of human iPSC generation. *Cell Stem Cell* **3:** 475-9.

Zwaka TP, Thomson JA (2005). A germ cell origin of embryonic stem cells? *Development* **132:** 227-33.

APPENDIX